

STIC-Biotech/ChemLib

174942

From: Ramirez, Delia
Sent: Thursday, December 22, 2005 3:53 PM
To: STIC-Biotech/ChemLib
Subject: 10/602220

Hi,

I would like to request the following searches:

1. SEQ ID NO: 15, 16 and 29 in the nucleic acid databases (commercial and interference)
2. an alignment of SEQ ID NO:16 and 29
3. SEQ ID NO:16 and 29 in the protein databases (commercial & interference)

Thank you very much,

Delia M. Ramirez, Ph.D.
Patent Examiner
Recombinant Enzymes-Art Unit 1652
USPTO
400 Dulany Street, Remsen Bldg., 2D74, Mail room 2C70
Alexandria, VA 22314
(571) 272-0938
delia.ramirez@uspto.gov

DEC 23 2005
USPTO

Searcher: _____
Searcher Phone: _____
Date Searcher Picked up: _____
Date completed: 4/3/06
Searcher Prep Time: _____
Online Time: _____

Type of Search
NA# _____ AA# _____
S/L: _____ Oligomer: _____
Encode/Transl: _____
Structure #: _____ Text: _____
Inventor: _____ Litigation: _____

Vendors and cost where applicable
STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other (Specify): _____

This Page Blank (uspto)

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1254	100.0	1254	6	AR653698	AR653698	Sequence
2	1254	100.0	1254	6	AR654966	AR654966	Sequence
3	1252.4	99.9	1266	6	AR653696	AR653696	Sequence
4	1252.4	99.9	1266	6	AR654964	AR654964	Sequence
5	1252.4	99.9	1278	6	AR653694	AR653694	Sequence
6	1252.4	99.9	1278	6	AR654962	AR654962	Sequence
7	1252.4	99.9	1284	6	AR653697	AR653697	Sequence
8	1252.4	99.9	1284	6	AR654965	AR654965	Sequence
9	1252.4	99.9	1230	6	CS135670	CS135670	Sequence
10	1252.4	99.9	1290	6	AR653692	AR653692	Sequence
11	1252.4	99.9	1290	6	AR654960	AR654960	Sequence
12	1252.4	99.9	1290	8	BT006864	BT006864	Homo sapi
13	1252.4	99.9	1290	11	AY889631	AY889631	Synthetic
14	1252.4	99.9	1290	11	AY889632	AY889632	Synthetic
15	1252.4	99.9	1290	11	AY892090	AY892090	Synthetic
16	1252.4	99.9	1290	11	BT007835	BT007835	Synthetic
17	1252.4	99.9	1296	6	AR653695	AR653695	Sequence
18	1252.4	99.9	1296	6	AR654963	AR654963	Sequence

241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGAGTGTGGATGGCTCCCAAAGA 300
241 TGAAGGATGAGGTTATGAGTACCTCTGCAATGATGAGTGTGGATGGCTCCCAAAGA 300
301 GATTTCAGAGGAGAGCTTCAGGCAGACCTCTGAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
301 GATTTCAGAGGAGAGCTTCAGGCAGACCTCTGAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
361 GCTAATATTATTTACACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
361 GCTAATATTATTTACACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
421 ACTGCGGAGGCTTCCTCGGAGTTTGTGATACGATACGATTCGATGATGATGATGATGATGAT 480
421 ACTGCGGAGGCTTCCTCGGAGTTTGTGATACGATTCGATGATGATGATGATGATGATGATGAT 480
481 GACTGGGAGTAGATCTGCTGCAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
481 GACTGGGAGTAGATCTGCTGCAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTGTTGATC 600
541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTGTTGATC 600
601 TCTGTGAGTGGCTCTTTATATGTTGAGGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660
601 TCTGTGAGTGGCTCTTTATATGTTGAGGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCAGTGGGAAATTTGCTGCAATTTGATGATTCCTGGAATAAGTATAAAG 720
661 CAGTACTGCAATCAGTGGGAAATTTGCTGCAATTTGATGATTCCTGGAATAAGTATAAAG 720
721 AGTATCTTGAATGAGCATCTTTTAAACAGGAGAAATTTGATGATGATGATGATGATGATGATGAT 780
721 AGTATCTTGAATGAGCATCTTTTAAACAGGAGAAATTTGATGATGATGATGATGATGATGATGAT 780
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAATTTGGCCCTCAGCTGGAATCAGCAA 840
781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAATTTGGCCCTCAGCTGGAATCAGCAA 840
841 GTAACCTCAGATGAGCTCTGGGCTATCATGCTGCTCTCTTTATTTATGCTTAATGACCTC 900
841 GTAACCTCAGATGAGCTCTGGGCTATCATGCTGCTCTCTTTATTTATGCTTAATGACCTC 900
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 960
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATGAGGATGAGGATGAGGATGAGGATGAGGAT 960
961 CAGGACCCCTTGGGCAAGGAGGTACCGCTTTAGACGGGAGACAACTTTGAAGTGTGG 1020
961 CAGGACCCCTTGGGCAAGGAGGTACCGCTTTAGACGGGAGACAACTTTGAAGTGTGG 1020
1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
1081 GGACCTCGCTCTTTATACCATCGAGTTGCTTCCCTGGGTTAAAGAGTGGCTGTATTCCT 1140
1081 GGACCTCGCTCTTTATACCATCGAGTTGCTTCCCTGGGTTAAAGAGTGGCTGTATTCCT 1140
1141 GCCTGCTTCATCAGAGCTCTCTCTGTTGAAAGGAGTGGCTGTATTCCTGATGATGATGATGATGATGAT 1200
1141 GCCTGCTTCATCAGAGCTCTCTCTGTTGAAAGGAGTGGCTGTATTCCTGATGATGATGATGATGATGAT 1200
1201 TCAAGGTTAAGAGTCAATATAATCCCAAGGAGTGTGTTTGTCTTCAAGCTATAA 1254
1201 TCAAGGTTAAGAGTCAATATAATCCCAAGGAGTGTGTTTGTCTTCAAGCTATAA 1254

RESULT 2
AR654966
LOCUS 1254 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 15 from patent US 6890748.

AR654966
AR654966.1 GI:67586505
Unknown.
ORGANISM
REFERENCE
1 (bases 1 to 1254)
Unclassified.
AUTHORS
Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE
Production of lysosomal enzymes in plants by transient expression
JOURNAL
Patent: US 6890748-A 15 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
1..1254
source
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 1254; DB 6; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAAGTCAAGGAGCCAGAACTACATCTGGGCTGAGCACTGGCAATGGATTGGCAAGAGCGCT 60
Db 1 ATGCAAGTCAAGGAGCCAGAACTACATCTGGGCTGAGCACTGGCAATGGATTGGCAAGAGCGCT 60
QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 181 GATTCTGCACTGAGGAGAACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGGC 240
Db 181 GATTCTGCACTGAGGAGAACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGGC 240
QY 241 TGGAGGATGCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGGAGGATGCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GATTTCAGAGGAGAGCTTCAGGCAGACCTTCAGGCAGACCTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGAGAGCTTCAGGCAGACCTTCAGGCAGACCTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATATTATTTACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
Db 361 GCTAATATTATTTACAGCAAGGAGCTGAAGCTAGGATTTATGAGATGTTGGAAATAAA 420
QY 421 ACTGCGGAGGCTTCCTCGGAGTTTGTGATACGATTCGATGATGATGATGATGATGATGATGATGAT 480
Db 421 ACTGCGGAGGCTTCCTCGGAGTTTGTGATACGATTCGATGATGATGATGATGATGATGATGATGAT 480
QY 481 GACTGGGAGTAGATCTGCTGCAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
Db 481 GACTGGGAGTAGATCTGCTGCAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 540
541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTGTTGATC 600
541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTGTTGATC 600
601 TCTGTGAGTGGCTCTTTATATGTTGAGGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660
601 TCTGTGAGTGGCTCTTTATATGTTGAGGCTTTCAGAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCAGTGGGAAATTTGCTGCAATTTGATGATTCCTGGAATAAGTATAAAG 720
661 CAGTACTGCAATCAGTGGGAAATTTGCTGCAATTTGATGATTCCTGGAATAAGTATAAAG 720
721 AGTATCTTGAATGAGCATCTTTTAAACAGGAGAAATTTGATGATGATGATGATGATGATGATGATGAT 780
721 AGTATCTTGAATGAGCATCTTTTAAACAGGAGAAATTTGATGATGATGATGATGATGATGATGATGAT 780

```
QY 781 GGTGGATGACCCAGATATGTTAGTGATTTGGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGATGACCCAGATATGTTAGTGATTTGGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTCATGTCTATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTCATGTCTATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGCAAGCAAGGTAACAGCTCTAGCAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGCAAGCAAGGTAACAGCTCTAGCAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTACGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTACGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
QY 1141 GCCTGCTTATCACAAGCTCTCTCTCTGTAAGGAGTGGCTCTTAATGAGTGGT 1200
DB 1141 GCCTGCTTATCACAAGCTCTCTCTCTGTAAGGAGTGGCTCTTAATGAGTGGT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
DB 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254

RESULT 3
LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 11 from patent US 6887696.
ACCESSION AR653696
VERSION AR653696.1 GI:67584281
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
     source          Location/Qualifiers
     1..1266
     /organism="unknown"
     /mol_type="genomic DNA"

ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCCCTGGCC 60
QY 61 CTGCTTTCTCGGACATCTCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
DB 61 CTGCTTTCTCGGACATCTCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAAGGAGGAGGCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAAGGAGGAGGCA 180
QY 181 GATTCTGCATCAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCTGCATCAGTGAAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
```

RESULT 4
AR654964
LOCUS

AR654964 1266 bp DNA linear PAT 13-JUN-2005

```
QY 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
DB 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCCAAAGA 300
QY 301 GATTCAAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGACGCTA 360
DB 301 GATTCAAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGACGCTA 360
QY 361 GCTAAATTATGTTTCAACAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATAAA 420
DB 361 GCTAAATTATGTTTCAACAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATAAA 420
QY 421 ACCTGCGAGGCTTCCCTGGAGTTTGGAGTACTACGACATTCGATGCGCCAGACCTTTGCT 480
DB 421 ACCTGCGAGGCTTCCCTGGAGTTTGGAGTACTACGACATTCGATGCGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAATAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTTACTGTGACAGATTTGGAATAATTG 540
QY 541 GCAGATGGTTTAAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAACATTTGTGTAC 600
DB 541 GCAGATGGTTTAAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAACATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTTGGCCCTTTTCAAAAGCCCAATTTATACAGAAATCGA 660
QY 661 CAGTACTGCATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATTAAG 720
DB 661 CAGTACTGCATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGTATTAAG 720
QY 721 AGTATCTGCACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGTCACACAGGG 780
DB 721 AGTATCTGCACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGTCACACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCAACTTTTGGCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTTGGCAACTTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTTAGCTATGATAAACCAGGAGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTACGCTGGGCTTAGCTATGATAAACCAGGAGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
QY 1141 GCTGCTTTCATCACAAGCTCTCTCTGTAAGGAGTGGCTCTTAATGAGTGGT 1200
DB 1141 GCTGCTTTCATCACAAGCTCTCTCTGTAAGGAGTGGCTCTTAATGAGTGGT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
DB 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTCAGCTATATAA 1254
```

```
DEFINITION Sequence 11 from patent US 6890748.
ACCESSION AR654964
VERSION AR654964.1 GI:67586503
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1266)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 11 10-MAY-2005;
LARGE SCALE Biology Corporation; Vacaville, CA
FEATURES
source
1..1266
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCCGCTTCCCTGGCC 60
QY 61 CTGCTTTCTTGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCTTGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
DB 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
QY 241 TGGAAAGATGACGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
DB 241 TGGAAAGATGACGGTTATGAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 301 GATTGAGAGGACAGCTTCAGGAGACCCCTCAGCGCTTCTCATGGATGGCTCCAGCTA 360
DB 301 GATTGAGAGGACAGCTTCAGGAGACCCCTCAGCGCTTCTCATGGATGGCTCCAGCTA 360
QY 361 GCTAAATTATCTTCACAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTGGAAATAA 420
DB 361 GCTAAATTATCTTCACAGCAAGGACTGAAGCTAGGGAATTTATGAGATGTTGGAAATAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGATTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGGATTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTACGATCTGCTAAATTTGATGGTGTGTACGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGAGTACGATCTGCTAAATTTGATGGTGTGTACGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGGTTATAGACATGTCTCGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
DB 541 GCAGATGGTTATAGACATGTCTCGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 720
QY 721 AGTATCTTGACCTGGAATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780
DB 721 AGTATCTTGACCTGGAATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACCAAGG 780
```

```
QY 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACTCAGATGCGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
DB 841 GTAACTCAGATGCGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTTAGAA 1254
RESULT 5
LOCUS AR653694
DEFINITION Sequence 7 from patent US 6887696.
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1278)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
LARGE SCALE Biology Corporation; Vacaville, CA
FEATURES
source
1..1278
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGGCC 60
QY 61 CTGCTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
DB 61 CTGCTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCTTCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
```

```
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGCGAGAGCTCATGGTCTCAGRAGGC 240
Qy 241 TGAAGAGATGACAGGTATAGTAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Db 241 TGAAGAGATGACAGGTATAGTAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Qy 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 361 GCTAATATATGTTACAGCAAGAGACTGAAGCTAGGATTTATGAGATGTTGGAATATAA 420
Db 361 GCTAATATATGTTACAGCAAGAGACTGAAGCTAGGATTTATGAGATGTTGGAATATAA 420
Qy 421 ACCTGGCAGGCTTCCTGGAGGTTTGGATATCTAGCAATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCTGGAGGTTTGGATATCTAGCAATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTTATAGCACATGTCCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTTATAGCACATGTCCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Qy 601 TCCTGTAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGAGAAATTTGATGTTGCTGACACAGGG 780
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGAGAAATTTGATGTTGCTGACACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCAGTAAATGTCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGCAGTAAATGTCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACACAGCTTAGACAGGAGACAACTTTGAAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACACAGCTTAGACAGGAGACAACTTTGAAAGTGTG 1020
Qy 1021 GAAACGACCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGGACAGGATGGT 1080
Db 1021 GAAACGACCTCTCAGGCTTAGCCTGGCTGTAGCTATGATATAACCGGACAGGATGGT 1080
Qy 1081 GGAACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
Db 1081 GGAACCTCGCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
Qy 1141 GCCTGCTTTCATCACAGAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTTCATCACAGAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
```

RESULT 6
AR654962

LOCUS AR654962 1278 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 7 from patent US 6890748.
ACCESSION AR654962
VERSION AR654962.1 GI:67586501
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1278)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 7 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source 1..1278
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"

Query Match 99.9%; Score 1252.4; DB 6; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 61 CTGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGACCCCT 120
Db 61 CTGTTTCTCTGGGACATCCTCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGACCCCT 120
Qy 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACACCTTGCAGTGCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACACCTTGCAGTGCAGGAGAGCCA 180
Qy 181 GATTCCTGCATCAGTCAGAGAGCTCTTCATGGAGATGCGAGAGCTCATGTGCTCAGAGGC 240
Db 181 GATTCCTGCATCAGTCAGAGAGCTCTTCATGGAGATGCGAGAGCTCATGTGCTCAGAGGC 240
Qy 241 TGAAGGATGAGGTTATGATGATCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGAGGTTATGATGATCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTCAGAGGAGAGCTTCAGGAGACCTCAGCAGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGAGCTTCAGGAGACCTCAGCAGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 361 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAATGTTGGAATATAA 420
Db 361 GCTAATATGTTACAGCAAGAGCTGAAGCTAGGATTTATGCAATGTTGGAATATAA 420
Qy 421 ACCTGGCAGGCTTCCTGGAGTTCCTGGATGATGATGATGATGATGATGATGATGATGAT 480
Db 421 ACCTGGCAGGCTTCCTGGAGTTCCTGGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTTATAGCACATGTCCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTTATAGCACATGTCCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Qy 601 TCCTGTAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCATCACTGGCGAAATTTGCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGAGAAATTTGATGTTGCTGACACAGGG 780
```

```
Db 721 AGTATCTTGGACTGCACATCTTTTAAACAGGAGAGAAATTGTTGATGTTGCTGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGACGTAATTCGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGACGTAATTCGCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTAACCACTTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTAACCACTTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
QY 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAGGTTAAGAGTCACATATAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254
Db 1201 TCAGGTTAAGAGTCACATATAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254

RESULT 7
AR653697 LOCUS 1284 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1284)
AUTHORS Garger, S. J., Turpen, T. H. and Kumagai, M. H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
    Location/Qualifiers
        source
            1..1284
                /organism="unknown"
                /mol_type="genomic DNA"
ORIGIN
    Query Match 99.9%; Score 1252.4; DB 6; Length 1284;
    Best Local Similarity 99.9%; Pred. No. 0;
    Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGGCTTGCCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGGCTTGCCTTCCTGGCC 60
QY 61 CTGCTTTCTTGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 120
Db 61 CTGCTTTCTTGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGTGCTGCCAGGAGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGTGCTGCCAGGAGCCA 180
QY 181 GATTCTGCACTAGTGAGAGCTCTTTCATGGAGATGCGAGGCTCATGCTCAGAGGC 240
```

RESULT 8

```
181 GATTCTGCACTAGTGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
241 TGGAAAGGATCAGAGTTATGAGTACCTCTGCAATGATGATGATGCTGCTCCCAAAGA 300
241 TGGAAAGGATCAGAGTTATGAGTACCTCTGCAATGATGATGATGCTGCTCCCAAAGA 300
301 GATTTCAGAAAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
301 GATTTCAGAAAGGAGAGCTTCAGGACAGACCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
361 GCTAATATATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
361 GCTAATATATGTTTACAGCAAAAGGAGCTGAAGCTAGGATTTATGCGAGATGTTGGAATAAA 420
421 ACTCGGAGAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGATGCTGCAAGATTTGCT 480
421 ACTCGGAGAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGATGCTGCAAGATTTGCT 480
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
481 GACTGGGAGTAGATCTGCTAAAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAAATAGACCTGGCGCAAGACATTTGTGTAC 600
541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAAATAGACCTGGCGCAAGACATTTGTGTAC 600
601 TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
601 TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCAATTTGATGATGATGATGATGATGATGATGAT 720
661 CAGTACTGCAATCAGCTGGCGAAATTTGCTGCAATTTGATGATGATGATGATGATGATGATGAT 720
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTGTTGATGTTGCTGACAGGG 780
721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTGTTGATGTTGCTGACAGGG 780
781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGGCTCAGCTGGAATCAGCAA 840
781 GGTGGAATGACCCAGATATGTTAGTATGTCGAACTTTGGGCTCAGCTGGAATCAGCAA 840
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTCGCATCAAT 960
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTCGCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
1021 GAACGACCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
1141 GCCTGCTTCATCACACAGCTCTCTCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT 1200
1201 TCAGGTTAAGAGTCACATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
1201 TCAGGTTAAGAGTCACATATAATCCACAGGCACTGTTTGTCTCAGCTAGAA 1254
```


AR654965	AR654965	Sequence 13 from patent US 6890748.	1284 bp	DNA	linear	PAT 13-JUN-2005
LOCUS	AR654965	Sequence 13 from patent US 6890748.	1284 bp	DNA	linear	PAT 13-JUN-2005
DEFINITION	AR654965	Sequence 13 from patent US 6890748.	1284 bp	DNA	linear	PAT 13-JUN-2005
ACCESSION	AR654965.1	GI:67586504				
VERSION	AR654965.1	GI:67586504				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 1284)					
AUTHORS	Garger,S.J., Turpen,T.H. and Kumagai,M.H.					
TITLE	Production of lysosomal enzymes in plants by transient expression					
JOURNAL	Patent: US 6890748-A 13 10-MAY-2005;					
FEATURES	Large Scale Biology Corporation; Vacaville, CA					
source	Location/Qualifiers					
	1..1284					
	/organism="unknown"					
ORIGIN	/mol_type="genomic DNA"					
	Query Match	99.9%; Score 1252.4; DB 6; Length 1284;				
	Best Local Similarity	99.9%; Pred. No. 0;				
	Matches 1253; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
QY	1	ATCAGCTGAGGAACCCAGAACTACATCTGGCGTTCGCCTTCGCCCTTCCTTCCTTGCGC	60			
Db	1	ATCAGCTGAGGAACCCAGAACTACATCTGGCGTTCGCCTTCGCCCTTCCTTCCTTGCGC	60			
QY	61	CTCGTTTCTCGGACATCCCTGGGCTTAGAGCACTGGACAAATGGATTGGCAAGACGCCT	120			
Db	61	CTCGTTTCTCGGACATCCCTGGGCTTAGAGCACTGGACAAATGGATTGGCAAGACGCCT	120			
QY	121	ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAAGGAGGCCA	180			
Db	121	ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAAGGAGGCCA	180			
QY	181	GATTCTGCATCAGTGAGAAAGCTCTTCATCGGAGATGGCAGAGCTCATGTCTCAGAAGGC	240			
Db	181	GATTCTGCATCAGTGAGAAAGCTCTTCATCGGAGATGGCAGAGCTCATGTCTCAGAAGGC	240			
QY	241	TGGAAGGATGCAGTTATGATGTAAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300			
Db	241	TGGAAGGATGCAGTTATGATGTAAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300			
QY	301	GATTTCAGAAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA	360			
Db	301	GATTTCAGAAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA	360			
QY	361	GCTAATATATGTTCA CAGCAAAGCACTGAAGCTAGGAGATTTATCAGATGTTGAAAATAAA	420			
Db	361	GCTAATATATGTTCA CAGCAAAGCACTGAAGCTAGGAGATTTATCAGATGTTGAAAATAAA	420			
QY	421	ACCTGGCGAGGCTTCCTTGGAGTTTTGGATCTACGACATTAATGCCAGACCTTTGCT	480			
Db	421	ACCTGGCGAGGCTTCCTTGGAGTTTTGGATCTACGACATTAATGCCAGACCTTTGCT	480			
QY	481	GACTCGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTTTACTGTCA CAGTTTGGAAAATTTG	540			
Db	481	GACTCGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTTTACTGTCA CAGTTTGGAAAATTTG	540			
QY	541	GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAACATTTGTGTAC	600			
Db	541	GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAACATTTGTGTAC	600			
QY	601	TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA	660			
Db	601	TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA	660			
QY	661	CAGTACTGCAATCACCTGGCGAAATTTTGCTGACATTCATGATTCCTGGAAAAGCTATAAAG	720			
Db	661	CAGTACTGCAATCACCTGGCGAAATTTTGCTGACATTCATGATTCCTGGAAAAGCTATAAAG	720			
QY	721	AGTATCTTGAGCTGGACATCTTTTAAAC CAGGAGAGAAATTTGTTGATGTTGCTGACACAGGG	780			

Db	721	AGTATCTTGAGCTGGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGCACGAGG	780		
Qy	781	GGTTGGAATGACCCAGATATGTTAGTGATTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA	840		
Db	781	GGTTGGNATGACCCAGATATGTTAGTGATTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA	840		
Qy	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATAATGACCTC	900		
Db	841	GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATAATGACCTC	900		
Qy	901	CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGTTAATGCGCATCAAT	960		
Db	901	CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGTTAATGCGCATCAAT	960		
Qy	961	CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACCTTTGAAAGTGTGG	1020		
Db	961	CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAACCTTTGAAAGTGTGG	1020		
Qy	1021	GAACGACCTCTCTCAGGCTTAGCTCGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080		
Db	1021	GAACGACCTCTCTCAGGCTTAGCTCGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080		
Qy	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTTAAGGAGTGGCTGTATTCCT	1140		
Db	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTTAAGGAGTGGCTGTATTCCT	1140		
Qy	1141	GCTCGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT	1200		
Db	1141	GCTCGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATCGACT	1200		
Qy	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA	1254		
Db	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA	1254		
RESULT 9	CS135670	1290 bp	DNA	linear	PAT 09-AUG-2005
LOCUS	Sequence 1 from Patent WO2005069192.				
DEFINITION	CS135670				
ACCESSION	CS135670.1	GI:72056330			
VERSION					
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1				
AUTHORS	Garman,S.C.				
TITLE	Crystal structure of human a-galactosidase				
JOURNAL	Patent: WO 2005069192-A 1 28-JUL-2005; Transkaryotic Therapies, Inc. (US); Selden, Richard F. (US); Garbocz, David N. (US); Treco, Douglas A. (US); NIH/NIHID (US); Garman, Scott C. (US); Borowski, Marianne (US); Kinoshita, Carol M. (US)				
FEATURES	Location/Qualifiers				
source	1..1290				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
ORIGIN					
Query Match	99.9%;	Score 1252.4;	DB 6;	Length 1290;	
Best Local Similarity	99.9%;	Pred. No. 0;			
Matches 1253;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
Qy	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCTTCGCTTCCTGGCC	60		
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCTTCGCTTCCTGGCC	60		
Qy	61	CTCGTTTCTTGGGACATCCTCGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCTT	120		

```
Db 61 CTCGTTTCTGGGACATCCTCGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCGCT 120
Qy 121 ACCATGGGCTGCTGCATCGGAGCGCTTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180
Db 121 ACCATGGGCTGCTGCATCGGAGCGCTTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTTCTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTTCTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TCGAAGGATGCAGGTATCAGTACCTCTGCAATGATGACTGTTGGATGCTCCCAAGA 300
Db 241 TCGAAGGATGCAGGTATCAGTACCTCTGCAATGATGACTGTTGGATGCTCCCAAGA 300
Qy 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATTATGTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAAGATGTTGGAATAAA 420
Qy 421 ACTCGCGCAGGCTTCCCTGGAGTTTGGATACCTACGACATTCGATGCCAGACCTTTGCT 480
Db 421 ACTCGCGCAGGCTTCCCTGGAGTTTGGATACCTACGACATTCGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGACTGCTTAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
Db 481 GACTGGGAGTAGACTGCTTAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
Qy 541 GCAGATGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTATC 600
Db 541 GCAGATGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTATC 600
Qy 601 TCGTGTAGTGGCTCTTTATATGTGCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCGTGTAGTGGCTCTTTATATGTGCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGTGTGCAATTCATGATGATTCCTGGAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGTGTGCAATTCATGATGATTCCTGGAAAGTATAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGAGAAATTTGATGTTGCTGACACAGGG 780
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGAGAAATTTGATGTTGCTGACACAGGG 780
Qy 781 GGTGGAAATCACCAAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATCACCAAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
Qy 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATAGCCCTCAAGCAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGCAAGCAAGGTTACAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGCAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAAACCGGCAGGAGATGGT 1080
Qy 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAAGGAGAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTTCATCACAGCTCTCTCCCTGTGAAAAGGAGAGCTAGGGTTCTATGAATGGACT 1200
```

```
Qy 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTCTTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATAATCCACAGGCACTGTTTCTTCTCAGCTAGAA 1254

RESULT 10
LOCUS AR653692 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6887696.
ACCESSION AR653692
VERSION AR653692.1 GI:67584277
KEYWORDS .
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITILE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 3 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source
1..1290
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 99.9%; Score 1252.4; DB 6; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
Qy 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGCGCCT 120
Db 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTTGGCAAGCGCCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTCAGCTGCCAGGAAGGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTCAGCTGCCAGGAAGGCCA 180
Qy 181 GATTCTTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCTTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGAAGGATGCAGGTATGATGATACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAGA 300
Db 241 TGAAGGATGCAGGTATGATGATACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTTCAGAAAGGAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTATCTTCACAGCAAGGAGCTCAAGCTAGGGAATTTATGCAAGATGTTGGAATAAA 420
Db 361 GCTAATTATCTTCACAGCAAGGAGCTCAAGCTAGGGAATTTATGCAAGATGTTGGAATAAA 420
Qy 421 ACCTGGCGCAGGCTTCCCTGGGAGTTTGGGATACCTACGACATTCGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCGCAGGCTTCCCTGGGAGTTTGGGATACCTACGACATTCGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGACTGCTTAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
Db 481 GACTGGGAGTAGACTGCTTAAATTTGATGTTGTTACTGTGACAGATTTGGAATAATTG 540
Qy 541 GCAGATGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTATC 600
Db 541 GCAGATGTTTAAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTATC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTGCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTGCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
```



```
Db 601 TCCTGTGAGTGGCCCTCTTTATATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCATCATCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATTTGCTGGACGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATTTGCTGGACGAGG 780
Qy 781 GGTGGAAATGACCCAGACATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGACATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTATGTCTTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTATGTCTTAATGACCTC 900
Qy 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Db 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTTGGCAAGCAAGGTACACGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTTGGCAAGCAAGGTACACGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
Qy 1081 GGAAGCTGCTTTATACATCGAGTGGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAAGCTGCTTTATACATCGAGTGGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTTGTCTCAGCTATAA 1254
```

```
RESULT 11
AR654960
LOCUS AR654960 1290 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 3 from patent US 6890748.
ACCESSION AR654960
VERSION AR654960.1 GI:67586497
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Garger, S.J., Turpen, T.H. and Kunagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 3 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1290
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 99.98; Score 1252.4; DB 6; Length 1290;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCCTGGCC 60
Qy 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 1200
```

```
Db 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGCTGCCAGGAAGAGCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGCTGCCAGGAAGAGCA 180
Qy 181 GATTCTTCGATCAGTGAAGAGCTCTTCATGAGAGATGCGAGCTCATGTGCTCAGAGAGC 240
Db 181 GATTCTTCGATCAGTGAAGAGCTCTTCATGAGAGATGCGAGCTCATGTGCTCAGAGAGC 240
Qy 241 TGGAGGATGCAAGTTATGATGACCTCTGATGATGATGATGATGATGATGATGATGATGATG 300
Db 241 TGGAGGATGCAAGTTATGATGACCTCTGATGATGATGATGATGATGATGATGATGATGATG 300
Qy 301 GATTTCAGAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Qy 361 GCTAATTTATGTTCAAGCAAGGACTGAACTAGGAGTTTATGCAATGTTGGAAATAAA 420
Db 361 GCTAATTTATGTTCAAGCAAGGACTGAACTAGGAGTTTATGCAATGTTGGAAATAAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTAGCAGCTTTCCTCATGGGATTCGCGAGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTAGCAGCTTTCCTCATGGGATTCGCGAGCT 480
Qy 481 GACTGGGAGTACATCTGCTAAAATTTGATGGTGTGTTTACCTGACAGCTTTGGAAAATTTG 540
Db 481 GACTGGGAGTACATCTGCTAAAATTTGATGGTGTGTTTACCTGACAGCTTTGGAAAATTTG 540
Qy 541 GCAGATGTTTAAAGCACAATGCTTGGCCCTGAAATAGGACTGCGCAGAGCAATGTTGTGAC 600
Db 541 GCAGATGTTTAAAGCACAATGCTTGGCCCTGAAATAGGACTGCGCAGAGCAATGTTGTGAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGAGGCTTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGAGGCTTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCATCACTGGCGAAATTTTGTGCAATGATGATTCCTCGAAAAGTATAAG 720
Db 661 CAGTACTGCATCACTGGCGAAATTTTGTGCAATGATGATTCCTCGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATGATTCCTGGAATGATG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATGATTCCTGGAATGATG 780
Qy 781 GGTGGAAATGACCCAGACATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGACATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTATGTCTTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTATGTCTTAATGACCTC 900
Qy 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Db 901 CGACATCATGACCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTTGGCAAGCAAGGTACACGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTTGGCAAGCAAGGTACACGCTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATATAAACCGGACGAGATTGGT 1080
Qy 1081 GGAAGCTGCTTTATACATCGAGTGGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAAGCTGCTTTATACATCGAGTGGCTTCCCTGGGTAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCAATCACAGCTCTCTCCCTGTGAAGGAAGCTAGGTTCTATGAATGACCT 1200
```


Db 1081 GGACCTCGCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTATTCCT 1140

Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGAGAGCTAGGGTTCTATGAATGGACT 1200

Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGAGAGCTAGGGTTCTATGAATGGACT 1200

Qy 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254

Db 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254

RESULT 13

LOCUS AY889631

DEFINITION Synthetic construct Homo sapiens clone FLH025769.01X galactosidase alpha (GLA) mRNA, complete cds.

ACCESSION AY889631

VERSION AY889631.1 GI:60655716

KEYWORDS Human ORF Project.

SOURCE synthetic construct

ORGANISM other sequences; artificial sequences.

REFERENCE 1 (bases 1 to 1290)

AUTHORS Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and LaBaer, J.

TITLE Cloning of human full-length CDS in Creator (TM) recombinational vector system

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1290)

AUTHORS Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and LaBaer, J.

TITLE Direct Submission

JOURNAL Submitted (04-JAN-2005) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141, USA

COMMENT This CDS clone is a part of a collection of human full-length expression clones generated by Harvard Institute of Proteomics. This ORF clone has been cloned with normalized stop-codon. The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-Dual vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence. Each clone is clonally isolated and full-length sequence-verified.

FEATURES

Source 1..1290

/organism="synthetic construct"

/mol_type="mRNA"

/db_xref="taxon:32630"

/clone="FLH025769.01X"

/lab_host="Escherichia coli DH5alpha T1 resistant"

/note="derived from MGC template"

gene 1..1290

/genes="GLA"

1..1290

/gene="GLA"

/codon_start=1

/transl_table=11

/product="galactosidase alpha"

/protein_id="AA32422.1"

/db_xref="GI:60655717"

/translation="MOLRNPEHLGCLALRFLALVSDIPGARLDNGLARTPTMGW LHWRFMNCNDQEEPDSCISEKLFMEAEIWMVSEGMKDGVEYLICIDCWMAPOKDS EGRLOADPQRPFGHIGQLANYVHSGKLGIYADVGNKTKCGFPGSPFGYDIAQTPA DNGVDLTKDGCCTDSEMLADGKMSLALNRGSRIVTSCEWPDYIMPPFPKNTTE IRQYCNHWRNPADIDDSWKSISILDWTSFNQERIVDVAGPGNDPDLVIGNFGLS WNOQVTOMALWAIMAAPLWMSNDRHISPOKALLQDKDVIADNPDLQKQVLEQD DNPVWFRPLSLAWAVAMINROEIGGPRSYTTAVASLGKGVACNPAFCITQLLPVKR KLGFEYWTSLRSHINPTGTVLLQLENTWQMSLKDLL"

ORIGIN

Query Match 99.9%; Score 1252.4; DB 11; Length 1290;

Best Local Similarity 99.9%; Pred. No. 0; Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCTCGTCCGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCTCGTCCGCC 60

Qy 61 CTGCTTCTTCGGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120

Db 61 CTGCTTCTTCGGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120

Qy 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGGCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGGCCA 180

Qy 181 GATTCTTCGCATCAGTACAGAACTCTTCATGGAGAGCTCTTCATGGAGAGTGGCAGAGGC 240

Db 181 GATTCTTCGCATCAGTACAGAACTCTTCATGGAGAGTGGCAGAGTTCATGGAGGC 240

Qy 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300

Db 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300

Qy 301 GATTCAAGAGGCAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGATTCGCAGCTA 360

Db 301 GATTCAAGAGGCAGACTTCAGGAGAGCCCTCAGGCGCTTTCCTCATGGATTCGCAGCTA 360

Qy 361 GCTAAATTATGTTTCAACAGCAAGGACTGAACTAGGAGTATTCAGAGATGTTGAAATAA 420

Db 361 GCTAAATTATGTTTCAACAGCAAGGACTGAACTAGGAGTATTCAGAGATGTTGAAATAA 420

Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTATTCAGGAGTATTCAGGAGTATTCAGGAGTATTC 480

Db 421 ACCTGCGCAGGCTTCCCTGGGAGTATTCAGGAGTATTCAGGAGTATTCAGGAGTATTC 480

Qy 481 GACTGGGAGTATGCTGCTAAATTTGATGTTGTTTCTGTCACAGTTCGGAATAATTTG 540

Db 481 GACTGGGAGTATGCTGCTAAATTTGATGTTGTTTCTGTCACAGTTCGGAATAATTTG 540

Qy 541 GCAGATGGTTTATAAGCACATGCTCTTGGGCTGAAATAGGACTGGCAGAGCATTTGTAC 600

Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGGCTGAAATAGGACTGGCAGAGCATTTGTAC 600

Qy 601 TCTGTGAGTGGCTCTTTATATGTCGCGCTTTCAAAGGCCCAATATATACAGAAATCCGA 660

Db 601 TCTGTGAGTGGCTCTTTATATGTCGCGCTTTCAAAGGCCCAATATATACAGAAATCCGA 660

Qy 661 CAGTACTGCAATCAGTGGGCAATTTTGTGACATTTGATGATTCCTGGAAAGTATTAAG 720

Db 661 CAGTACTGCAATCAGTGGGCAATTTTGTGACATTTGATGATTCCTGGAAAGTATTAAG 720

Qy 721 AGTATCTTGGACTTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACGGG 780

Db 721 AGTATCTTGGACTTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACGGG 780

Qy 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGGCTTCAGTGGAAATCAGCAA 840

Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGGCTTCAGTGGAAATCAGCAA 840

Qy 841 GTAATCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

Db 841 GTAATCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGCTAAATGGCATCAAT 960

Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGCTAAATGGCATCAAT 960

Qy 961 CAGGACCCCTTGGGCAAGCAAGGCTACACAGCTTTAGACAGGAGCAACCTTTGAAGTGTGG 1020

Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACACAGCTTTAGACAGGAGCAACCTTTGAAGTGTGG 1020

Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCAGGAGAGATTGGT 1080

Db	1021	GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGGCGAGGAGTTGCT	1080	KLGFVWTSRLRSHINPTGTVLLQLLENTMQMSLKDLL"
Qy	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGCGCTGTATCTCT	1140	Query Match 99.9%; Score 1252.4; DB 11; Length 1290;
Db	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGCGCTGTATCTCT	1140	Best Local Similarity 99.9%; Pred. No. 0;
Qy	1141	GCTGCTTATACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT	1200	Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Db	1141	GCTGCTTATACACAGCTCTCTCCCTGTGAAAGGAAGCTAGGTTCTATGAATGGACT	1200	
Qy	1201	TCAGGTTAAGAGTACATATAATCCACAGGACCTGTTTGTCTCAGCTATAA	1254	
Db	1201	TCAGGTTAAGAGTACATATAATCCACAGGACCTGTTTGTCTCAGCTAGBA	1254	
RESULT 14				
LOCUS	AY889632			
DEFINITION	Synthetic construct Homo sapiens clone FLH025770.01X galactosidase alpha (GLA) mRNA, complete cds.			
ACCESSION	AY889632			
VERSION	AY889632.1	GI:60655718		
KEYWORDS	Human ORF Project.			
SOURCE	synthetic construct			
ORGANISM	other sequences; artificial sequences.			
REFERENCE	1 (bases 1 to 1290)			
AUTHORS	Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and Labaer, J.			
TITLE	Cloning of human full-length CDS in Creator (TM) recombinational vector system			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1290)			
AUTHORS	Hines, L., Rolfs, A., Jepson, D., Moreira, D., Raphael, J., Kelley, F., Shen, B., Halleck, A., Koundinya, M., Hu, Y., Zuo, D., Taycher, E., Williamson, J. and Labaer, J.			
TITLE	Direct Submission			
JOURNAL	Submitted (04-JAN-2005) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141, USA			
COMMENT	This CDS clone is a part of a collection of human full-length expression clones generated by Harvard Institute of Proteomics. This ORF clone has been cloned with normalized stop-codon. The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-Dual vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence. Each clone is clonally isolated and full-length sequence-verified.			
FEATURES	Location/Qualifiers			
source	1..1290			
	/organism="synthetic construct"			
	/mol_type="mRNA"			
	/db_xref="taxon:32630"			
	/clone="FLH025770.01X"			
	/lab_host="Escherichia coli DH5alpha T1 resistant"			
	/note="derived from MGC template"			
gene	1..1290			
	/gene="GLA"			
CDS	1..1290			
	/gene="GLA"			
	/codon_start=1			
	/transl_table=11			
	/product="galactosidase alpha"			
	/protein_id="AA32423.1"			
	/db_xref="GI:60655719"			
	/translation="MQLRNPDLHLGALALRFLALVSWDIPGARLDNGLRPTMGW LHWERWNCNDCQEEEDSCISEKLFMEMELMWSEGWKAGCYFLCDDCWAPQRDS EGRLODPQRFPHGIQRLANYVHSGKLGIGYADVGNKTCAGFPFGYDYIDAQTPA DWGVLLKFDGCTCDSENLADGYKMSLALNRGSIYVSCWPLMYMFPKPNYTE IROYCNHRNFADIDDSWKSISKILDWTSFNQERIYDVAFPGWNPDMILVIGNFGLS WNOQVTQMALWATMAAPLWMSNDRHI SPOKALLQDKVIAINQDPLGKQYLRQG DNFVWRPLSGLAWAVAMINRQIEIGGPRSYTIAVASLKGKGVACNPACFITQLPVKR			


```
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAAT 960
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1140
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1141 GCCTGTCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1141 GCCTGTCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATGAA 1254
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Search completed: December 31, 2005, 19:26:44
Job time : 6626 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 31, 2005, 09:31:46 ; Search time 843 Seconds
(without alignments)
9914.027 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1. atgcagctgagggaaccaga.....ctgtttgttcagctataa 1254

Scoring table: / IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues
Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_21.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001as.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002as.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003as.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004as.*
 - 13: Geneseqn2004bs.*
 - 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1254	100.0	1254	6	AAD45225 Human rga
2	1254	100.0	1254	10	ADD84754 Human alp
3	1254	100.0	1254	12	ADM48684 Human wil
4	1254	100.0	1254	13	ADU66919 Human alp
5	1254	100.0	1254	14	AEA27448 Human alp
6	1252.4	99.9	1266	6	AAD45223 Human rga
7	1252.4	99.9	1266	10	ADD84750 Human alp
8	1252.4	99.9	1266	12	ADJ88276 Human wil
9	1252.4	99.9	1266	12	ADM48680 Human wil
10	1252.4	99.9	1266	13	ADU66915 Human alp
11	1252.4	99.9	1266	14	AEA27444 Human alp
12	1252.4	99.9	1278	6	AAD45221 Human rga
13	1252.4	99.9	1278	10	ADD84746 Human alp
14	1252.4	99.9	1278	12	ADJ88272 Human rga
15	1252.4	99.9	1278	12	ADM48676 Human wil
16	1252.4	99.9	1278	13	ADU66911 Human alp
17	1252.4	99.9	1278	14	AEA27440 Human alp
18	1252.4	99.9	1284	6	AAD45224 Human rga
19	1252.4	99.9	1284	10	ADD84752 Human alp

20	1252.4	99.9	1284	12	ADJ88278 Human WT
21	1252.4	99.9	1284	12	ADM48682 Human wil
22	1252.4	99.9	1284	13	ADU66917 Human alp
23	1252.4	99.9	1284	14	AEA27446 Human alp
24	1252.4	99.9	1290	6	AAD45219 Human WT
25	1252.4	99.9	1290	10	ADD84742 Human alp
26	1252.4	99.9	1290	12	ADJ88268 Human WT
27	1252.4	99.9	1290	12	ADM48672 Human wil
28	1252.4	99.9	1290	12	ADN49739 Human alp
29	1252.4	99.9	1290	13	ADU74415 Human alp
30	1252.4	99.9	1290	13	ADU66907 Human alp
31	1252.4	99.9	1290	14	AEA27436 Human alp
32	1252.4	99.9	1290	14	AEA27436 Human alp
33	1252.4	99.9	1296	6	AAD45222 Human rga
34	1252.4	99.9	1296	10	ADD84748 Human alp
35	1252.4	99.9	1296	12	ADJ88274 Human WT
36	1252.4	99.9	1296	12	ADM48678 Human wil
37	1252.4	99.9	1296	13	ADU66913 Human alp
38	1252.4	99.9	1296	14	AEA27442 Human alp
39	1252.4	99.9	1306	8	ABZ79777 Human alp
40	1252.4	99.9	1308	6	AAD45220 Human WT
41	1252.4	99.9	1308	10	ADD84744 Human alp
42	1252.4	99.9	1308	12	ADJ88270 Human WT
43	1252.4	99.9	1308	12	ADM48674 Human wil
44	1252.4	99.9	1308	13	ADU66909 Human alp
45	1252.4	99.9	1308	14	AEA27438 Human alp

ALIGNMENTS

RESULT 1
AAD45225
ID AAD45225 standard; DNA; 1254 BP.
XX AC AAD45225;
XX 27-DEC-2002 (first entry)
XX Human rGAL-12 DNA.
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-12; gene; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..1254
XX /*tag= a
XX /product= "Human rGAL-12 protein"
XX US2002088024-A1.
XX 04-JUL-2002.
XX 13-NOV-2001; 2001US-00993059.
XX 26-JUL-2000; 2000US-00626127.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
XX P-PSDB; AAE28212.
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX storage diseases.
XX Claim 1; Page 47-49; 88pp; English.

CC The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-12 DNA
XX

SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 6; Length 1254;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCAGCTGAGNACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC	60
DB	1	ATGCAGCTGAGNACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTTCCTGGCC	60
QY	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGCAAAATGGATTGGCAAGAGCCCT	120
DB	61	CTCGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGCAAAATGGATTGGCAAGAGCCCT	120
QY	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
DB	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
QY	181	GATTCCTGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC	240
DB	181	GATTCCTGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC	240
QY	241	TGGAAGGATGCAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	300
DB	241	TGGAAGGATGCAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	300
QY	301	GATTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGGCTA	360
DB	301	GATTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGGCTA	360
QY	361	GCTAATATGTTTACAGCAAGAGCTGAGGAGTTTATGAGATGTTGAGATGTTGAGATG	420
DB	361	GCTAATATGTTTACAGCAAGAGCTGAGGAGTTTATGAGATGTTGAGATGTTGAGATG	420
QY	421	ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCATTGATGATGATGATGATGATG	480
DB	421	ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTAGCATTGATGATGATGATGATGATG	480
QY	481	GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTTACTGTGACAGATTGTTGTA	540
DB	481	GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTTACTGTGACAGATTGTTGTA	540
QY	541	GCAGATGTTTAAAGCAGATCTCTGGGCTGATAGGAGTGGCAGAGCATTTGTTGATC	600
DB	541	GCAGATGTTTAAAGCAGATCTCTGGGCTGATAGGAGTGGCAGAGCATTTGTTGATC	600
QY	601	TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA	660
DB	601	TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA	660
QY	661	CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATGATGATGATGATGATGATG	720
DB	661	CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATGATGATGATGATGATGATG	720
QY	721	AGTATCTGCACTGCAATCTTTTACCAGGAGAGATTTGATGATGATGATGATGATGATG	780
DB	721	AGTATCTGCACTGCAATCTTTTACCAGGAGAGATTTGATGATGATGATGATGATGATG	780
QY	781	GGTTGGAATCACCAGATATGTTAGTATGTTGGAACCTTTGGCCCTCAGCTGGAATCAGCA	840
DB	781	GGTTGGAATCACCAGATATGTTAGTATGTTGGAACCTTTGGCCCTCAGCTGGAATCAGCA	840

QY	841	CTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCAATGCTAATGACCTC	900
DB	841	CTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCAATGCTAATGACCTC	900
QY	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT	960
DB	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT	960
QY	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTG	1020
DB	961	CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTG	1020
QY	1021	GAGGACCTCTCTCAGGCTTAGCTGCTGCTATGATATAACCGGACGAGAGATTGGT	1080
DB	1021	GAGGACCTCTCTCAGGCTTAGCTGCTGCTATGATATAACCGGACGAGAGATTGGT	1080
QY	1081	GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTTGAATCCT	1140
DB	1081	GGACCTCGCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTTGAATCCT	1140
QY	1141	GCTGCTTATCAGCAGCTCTCCTCTGTAAGGAGAGTGGCTTCTATGATGAGT	1200
DB	1141	GCTGCTTATCAGCAGCTCTCCTCTGTAAGGAGAGTGGCTTCTATGATGAGT	1200
QY	1201	TCAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTATAA	1254
DB	1201	TCAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTATAA	1254

RESULT 2

ADD84754

ID ADD84754 standard; DNA; 1254 BP.

XX ADD84754;

XX 29-JAN-2004 (first entry)

XX Human alpha-galactosidase rGAL-12 DNA.

XX Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;

XX enzyme replacement therapy; lysosomal disease.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 1..1254

XX /*tag= a

XX /product= "Human rGAL-12"

XX US2003106095-A1.

XX 05-JUN-2003.

XX 20-MAR-2002; 2002US-00103327.

XX 26-JUL-2000; 2000US-00626127.

XX 13-NOV-2001; 2001US-00993059.

XX (GARG/) GARGER S J.

XX (TURP/) TURPEN T H.

XX (KUMA/) KUMAGAI M H.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2003-801257/75.

XX P-PSDB; ADD84755.

XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.

XX Claim 1; SEQ ID NO 15; 77pp; English.

XX The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents DNA encoding a human alpha-
 CC galactosidase derivative polypeptide of the invention.

XX SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 10; Length 1254;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60

QY 61 CTCGTTTCCTGGACATCTCTGGGCTAGAGCACTGGCAATGATTTGGCAAGACGCT 120
 DB 61 CTCGTTTCCTGGACATCTCTGGGCTAGAGCACTGGCAATGATTTGGCAAGACGCT 120

QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAGGCCA 180
 DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAGGCCA 180

QY 181 GATTTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 DB 181 GATTTCCTGTCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGGAAAGATCAGGTTATGATGATACCTCTGCATGATGATGATGATGATGATGATGATGAT 300
 DB 241 TGGAAAGATCAGGTTATGATGATACCTCTGCATGATGATGATGATGATGATGATGATGAT 300

QY 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGGCTTCTCATGGGATTCGCCAGCTA 360
 DB 301 GATTGAGAGGAGACTTTCAGGAGAGCCCTCAGGCTTCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAAATATGTTTACAGCAAGAGCTGAAGCTAGGAGATTTATGATGATGATGATGATGATGAT 420
 DB 361 GCTAAATATGTTTACAGCAAGAGCTGAAGCTAGGAGATTTATGATGATGATGATGATGATGAT 420

QY 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATATCTAGCAATGATGATGATGATGATGATGATGAT 480
 DB 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATATCTAGCAATGATGATGATGATGATGATGATGAT 480

QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
 DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 540

QY 541 GCAGATGGTTTATAGCACATGCTCTGGCCCTGCAATAGGACTGGCAGAGCAATGATGATGATGAT 600
 DB 541 GCAGATGGTTTATAGCACATGCTCTGGCCCTGCAATAGGACTGGCAGAGCAATGATGATGATGAT 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 DB 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGTGGGCAATTTTCTGACATGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 661 CAGTACTGCAATCAGTGGGCAATTTTCTGACATGATGATGATGATGATGATGATGATGATGATGAT 720

QY 721 AGTATCTTGAGCTGGACATCTTTTAAACAGAGAGAAATGCTGATGATGATGATGATGATGATGAT 780
 DB 721 AGTATCTTGAGCTGGACATCTTTTAAACAGAGAGAAATGCTGATGATGATGATGATGATGATGAT 780

QY 781 GGTGGAATGACCCAGATATGTTAGTATGGAATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
 DB 781 GGTGGAATGACCCAGATATGTTAGTATGGAATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCAATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCAATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900

DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960

DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960

QY 961 CAGGACCCCTTTGGGCAAGCAAGGTTACGAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020

DB 961 CAGGACCCCTTTGGGCAAGCAAGGTTACGAGCTTTAGACAGGGAGACAACTTTGAAGTGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTTAGCTCGGCTGTAGCTATGATAAACCGGCAAGGAGATTGGT 1080

DB 1021 GAACGACCTCTCTCAGGCTTTAGCTCGGCTGTAGCTATGATAAACCGGCAAGGAGATTGGT 1080

QY 1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTTAATCCT 1140

DB 1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTCTTAATCCT 1140

QY 1141 GCCTGCTTATACACAGCTCTCTCTGTGAAAGGAGAGCTAGGCTTCTATGATGAGACT 1200

DB 1141 GCCTGCTTATACACAGCTCTCTCTGTGAAAGGAGAGCTAGGCTTCTATGATGAGACT 1200

QY 1201 TCAAGGTTAAGAGTCAACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

DB 1201 TCAAGGTTAAGAGTCAACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 3
 ADM48684
 ID ADM48684 standard; DNA; 1254 BP.
 XX AC ADM48684;
 XX AC
 XX DT 03-JUN-2004 (first entry)
 XX DE Human wild type rGAL-12 DNA.
 XX KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
 XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
 XX Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
 XX gene; ds.
 XX OS Homo sapiens.
 XX Key Location/Qualifiers
 XX 1. .1254
 XX CDS /*tag= a
 XX FT /product= "GAL-12 protein"
 XX FT
 XX PN US2004023281-A1.
 XX PD 05-FEB-2004.
 XX PF 23-JUN-2003; 2003US-00602220.
 XX PR 26-FEB-1988; 88US-00160766.
 XX PR 26-FEB-1988; 88US-00160771.
 XX PR 17-FEB-1989; 89US-00310881.
 XX PR 22-OCT-1990; 90US-00600244.
 XX PR 31-JUL-1992; 92US-00923692.
 XX PR 30-DEC-1992; 92US-00997733.
 XX PR 29-DEC-1993; 93US-00176414.
 XX PR 19-JAN-1994; 94US-00184237.
 XX PR 14-OCT-1994; 94US-00324003.
 XX PR 21-MAY-1999; 99US-00316572.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 XX (KUMA/) KUMAGAI M H.
 XX (POGU/) POGUE G P.
 XX (ERWL/) ERWIN R L.
 XX (GRIL/) GRILL L K.

XX PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
 XX DR WPI; 2004-142650/14.
 XX P-PSDB; ADM48685.
 XX PT New alpha-galactosidase polypeptides, useful in producing recombinant
 PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
 PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
 PT disease.
 XX PS Disclosure; SEQ ID NO 15; 72pp; English.
 XX CC The present invention relates to novel galactosidase (Gal) proteins such
 CC as rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions
 CC of the present invention are useful for producing recombinant lysosomal
 CC enzymes for enzyme replacement therapy for treating human and animal
 CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
 CC human wild type rGAL-12 DNA used in the exemplification of the invention.
 XX SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 12; Length 1254;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
 DB 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
 QY 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGACGCT 120
 DB 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGGACGCT 120
 QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTCAGTCCAGGAGAGCCCA 180
 DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTCAGTCCAGGAGAGCCCA 180
 QY 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
 QY 241 TGAAGGATGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
 DB 241 TGAAGGATGCAGGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
 QY 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 DB 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
 QY 361 GCTAATATGTTCCACAGCAAGGAGCTGAAGCTAGGATTTATCGAGATGTTGAAATATA 420
 DB 361 GCTAATATGTTCCACAGCAAGGAGCTGAAGCTAGGATTTATCGAGATGTTGAAATATA 420
 QY 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACATGACATTCATGATGATGATGATGATGAT 480
 DB 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACATGACATTCATGATGATGATGATGATGAT 480
 QY 481 GACTGGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 DB 481 GACTGGGAGTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
 QY 541 GCAGATGGTTATAGCAGCATGTCTTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 DB 541 GCAGATGGTTATAGCAGCATGTCTTCCTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 QY 601 TCCCTGAGTGGCTCTTTATATGTCGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 DB 601 TCCCTGAGTGGCTCTTTATATGTCGCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACTGATGATGATGATGATGATGATGATGAT 720
 DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACTGATGATGATGATGATGATGATGATGAT 720

DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCACTGATGATGATGATGATGATGATGATGAT 720
 QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATCTTCTGACCAAGG 780
 DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATCTTCTGACCAAGG 780
 QY 781 GGTGTGAATGACCCAGATATGTTAGTGTGGCACTTTGGGCTCAGCTGGAATCAGCAA 840
 DB 781 GGTGTGAATGACCCAGATATGTTAGTGTGGCACTTTGGGCTCAGCTGGAATCAGCAA 840
 QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
 DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
 QY 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGAGGAGATGGT 1080
 DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGAGGAGATGGT 1080
 QY 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTAATCCT 1140
 DB 1081 GGACCTCGCTCTTATACCATCCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTAATCCT 1140
 QY 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 DB 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
 QY 1201 TCAAGTTAAGAGTCACATAAATCCACAGGACCTGTTTGGCTTCAGCTATAA 1254
 DB 1201 TCAAGTTAAGAGTCACATAAATCCACAGGACCTGTTTGGCTTCAGCTATAA 1254

RESULT 4
 ADU66919
 ID ADU66919 standard; DNA; 1254 BP.
 XX
 AC ADU66919;
 XX
 DT 10-FEB-2005 (first entry)
 XX
 DE Human alpha-galactosidase protein encoding DNA #7.
 XX
 KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease; Hurler's disease;
 KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT CDS 1..1254
 FT /*tag= a
 FT /product= "Human alpha-galactosidase protein"
 XX
 PN US2004234516-A1.
 XX
 PD 25-NOV-2004.
 XX
 PF 21-MAY-2004; 2004US-00851388.
 XX
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-0093059.
 PR 20-MAR-2002; 2002US-00103327.
 XX
 PA (LARG-) LARGE SCALE BIOLOGY CORP.
 XX

PI Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
DR P-PSDB; ADU66920.
XX
PT A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
PS Disclosure; SEQ ID NO 15; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein encoding DNA.
XX
SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 13; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCCTGGCC 60

QY 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGCGCCT 120
DB 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATGGCAAGCGCCT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180

QY 181 GATTCTGTCATCAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCTGTCATCAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 241 TGAAGGATCGAGGTATAGTACCTCTGATGATGACCTGTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATCGAGGTATAGTACCTCTGATGATGACCTGTGGATGGCTCCCAAGA 300

QY 301 GATTGAGAGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTGAGAGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360

QY 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATATAA 420
DB 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATATAA 420

QY 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480
DB 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACCTACGACATTCATGCCAGACTTTGCT 480

QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540

QY 541 GCAGATGGTTAAGCAGCATGTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTATC 600
DB 541 GCAGATGGTTAAGCAGCATGTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTATC 600

QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCGACATTGATGATTCCTCGAAAAGTATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCGACATTGATGATTCCTCGAAAAGTATAAG 720

QY 721 AGTATCTTGACATGCGACATCTTTTAACCCAGAGAGATTTGATGTTGCTGGACCGGG 780
DB 721 AGTATCTTGACATGCGACATCTTTTAACCCAGAGAGATTTGATGTTGCTGGACCGGG 780

QY 781 GGTGGAATCACCAGAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATCACCAGAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
DB 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900

QY 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCCATCAAT 960
DB 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATATAAACCCGCGAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTTAGCTGGGCTGTAGCTATGATATAAACCCGCGAGGAGATTGGT 1080

QY 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTTAAGAGTGGCTGTAAATCCT 1140
DB 1081 GGAACCTGCTCTTATACCATCGCAGTTGCTTCCCTGGGTTAAGAGTGGCTGTAAATCCT 1140

QY 1141 GCCTGCTTCATACACAGCTCCCTCTGTAAGGAAGCTAGGGTTCTATGATGAGTACT 1200
DB 1141 GCCTGCTTCATACACAGCTCCCTCTGTAAGGAAGCTAGGGTTCTATGATGAGTACT 1200

QY 1201 TCAAGGTTAAGAGTCACTATAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCACTATAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254

RESULT 5
AEA27448
ID AEA27448 standard; DNA; 1254 BP.
XX
AC AEA27448;
XX
DT 11-AUG-2005 (first entry)
XX
DE Human alpha-galactosidase DNA, rGAL-12, SEQ ID NO: 15.
XX
KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1..1254
FT CDS /*tag= a
FT /product= "Human alpha-galactosidase protein"
XX
PN US2005125859-A1.
XX
PD 09-JUN-2005.
XX
PF 08-NOV-2004; 2004US-00984389.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.

XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX PA
XX PI
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI: 2005-404004/41.
DR P-P8DB; AEA27449.
XX
PT New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
PS Disclosure; SEQ ID NO 15; 88pp; English.
XX
CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase DNA.
XX
SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Query Match 100.0%; Score 1254; DB 14; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCGCTTCGCTTCCTGGCC 60

QY 61 CTCGTTTCTTGGACATCCTCGGGCTTAGAGCACTGGAACAATGGATGGCAAGGACGCT 120
DB 61 CTCGTTTCTTGGACATCCTCGGGCTTAGAGCACTGGAACAATGGATGGCAAGGACGCT 120

QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATCTGCAACCTTCAGCTCCAGGAGAGGCCA 180
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATCTGCAACCTTCAGCTCCAGGAGAGGCCA 180

QY 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
DB 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240

QY 241 TGAAGAGATCAGGTTTATGATGACCTCTGCAATGATGACTGTGTTGGATGGCTCCCAAGA 300
DB 241 TGAAGAGATCAGGTTTATGATGACCTCTGCAATGATGACTGTGTTGGATGGCTCCCAAGA 300

QY 301 GATTGAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTCCTCATGGGATTCGCGAGCTA 360
DB 301 GATTGAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTCCTCATGGGATTCGCGAGCTA 360

QY 361 GCTAATTTATCTTCACAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
DB 361 GCTAATTTATCTTCACAGCAAGGAGCTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420

QY 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATGATGACATGATGATGATGATGATGATGATGAT 480
DB 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATGATGACATGATGATGATGATGATGATGATGAT 480

QY 481 GACTGGGAGTACATCTGCTAATATTTGATGTTGTTTACTGTGACAGTTTGGAAAATTG 540
DB 481 GACTGGGAGTACATCTGCTAATATTTGATGTTGTTTACTGTGACAGTTTGGAAAATTG 540

QY 541 GCAGATGTTTATAGACATGCTCTTCGCGCTTGAATAGGAGCTGGCAGAGCAATGTTGTAC 600
DB 541 GCAGATGTTTATAGACATGCTCTTCGCGCTTGAATAGGAGCTGGCAGAGCAATGTTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGGCCAAATATATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGGCCAAATATATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGTGACATTTGATGATTTCTGGAAAAGATATAAG 720
DB 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGTGACATTTGATGATTTCTCTGGAAAAGATATAAG 720

QY 721 AGTATCTTGGACTGGACATCTTTTAACACAGGAGAGAAATTTGATGATTTCTGTCGACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAACACAGGAGAGAAATTTGATGATTTCTGTCGACAGGG 780

QY 781 GGTGGAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAAGCTTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATTAAGCTTC 900

QY 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGCTTAATTGCCATCAAT 960
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGCTTAATTGCCATCAAT 960

QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACGAGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGGACGAGATTTGGT 1080

QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTAATCCT 1140

QY 1141 GCCTGCTTCATACACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200
DB 1141 GCCTGCTTCATACACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTCTTATGAATGGACT 1200

QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254

RESULT 6
AAD45223
ID AAD45223 standard; DNA; 1266 BP.
XX
AC AAD45223;
XX 27-DEC-2002 (first entry)
DT
XX Human rGAL-8 DNA.
DE
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-8; gene; ds.
KW
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1266
XX FT /*tag= a
XX FT /product= "Human rGAL-8 protein"
XX
XX US2002088024-A1.
XX
XX 04-JUL-2002.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX

XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2002-681656/73.
DR P-PSDB; AAE28210.
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
PT storage diseases.
XX
XX Claim 1; Page 42-44; 88pp; English.
XX The invention relates to human alpha-galactosidase truncated at the
CC carboxy terminus and the production of enzymatically active recombinant
CC human and animal lysosomal enzymes. The invention is useful for producing
CC lysosomal enzymes for treating lysosomal storage diseases, producing
CC altered or mutated proteins, enzymatically active or otherwise, to serve
CC as precursors or substrates for further in vivo or in vitro processing to
CC a specialised industrial form for research or therapeutic uses, to
CC produce more effective therapeutic enzyme, for producing antibodies
CC against lysosomal enzymes for medical diagnostic use, and in any
CC commercial process that involves substrate hydrolysis. The present
CC sequence is human rGAL-8 DNA
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
Query Match 99.9%; Score 1252.4; DB 6; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 61 CTGCTTCTTGGGACATCTCTGGGCTAGAGCACTGGCAATGGATGGCAAGAGCGCT 120
DB 61 CTGCTTCTTGGGACATCTCTGGGCTAGAGCACTGGCAATGGATGGCAAGAGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATCTGCAACCTTCACTGCCAGGAGGCCA 180
QY 181 GATTCTGTGATCAGTGAGAGCTCTTCATGAGAGTGGCAGAGCTCATGTGCTCAGAAGGC 240
DB 181 GATTCTGTGATCAGTGAGAGCTCTTCATGAGAGTGGCAGAGCTCATGTGCTCAGAAGGC 240
QY 241 TGAAGAGATCAGGTTATGATGATACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
DB 241 TGAAGAGATCAGGTTATGATGATACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTGAGAGGCTTCAAGGAGACCTTCAAGGAGCTTCTTCATGGATGGCTCCCAAGTA 360
DB 301 GATTGAGAGGCTTCAAGGAGACCTTCAAGGAGCTTCTTCATGGATGGCTCCCAAGTA 360
QY 361 GCTAATTATGTTTCAAGCAAGGAGCTGAAGCTAGGAGTTTATGAGATGTTGAAATATA 420
DB 361 GCTAATTATGTTTCAAGCAAGGAGCTGAAGCTAGGAGTTTATGAGATGTTGAAATATA 420
QY 421 ACCTGGCGAGGCTTCCCTGGAGTTTGGATACCTAGCAGATTTGATGCCAGACTTTGGT 480
DB 421 ACCTGGCGAGGCTTCCCTGGAGTTTGGATACCTAGCAGATTTGATGCCAGACTTTGGT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTGTCTACTGTGACAGTTTGGAAATTTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTGTCTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGGTTATAGCAGCATGCTCTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600
DB 541 GCAGATGGTTATAGCAGCATGCTCTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
QY 721 AGTATCTTGGACTGGGACATCTTTTAAACCAGGAGAGAAATTTGATGATTTGATGACCAAGG 780
DB 721 AGTATCTTGGACTGGGACATCTTTTAAACCAGGAGAGAAATTTGATGATTTGATGACCAAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTATACACAGCTCTCTCTGTGTAAGAGAGCTAGGTTCTATGATGAGCT 1200
DB 1141 GCCTGCTTATACACAGCTCTCTCTGTGTAAGAGAGCTAGGTTCTATGATGAGCT 1200
QY 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
RESULT 7
ADD84750
ID ADD84750 standard; DNA; 1266 BP.
XX
AC ADD84750;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human alpha-galactosidase rGAL-8 DNA.
XX
KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 1..1266
FT /*tag= a
FT /product= "Human rGAL-8"
XX
PN US2003106095-A1.
XX
PD 05-JUN-2003.
XX
PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.

```
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
PI WPI; 2003-801257/75.
XX P-PSDB; ADD84751.
DR New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX Claim 1; SEQ ID NO 11; 77pp; English.
XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
Query Match 99.9%; Score 1252.4; DB 10; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGNACCAGACATACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 61 CTCGTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGCAATGGATTGGCAAGAGCCCT 120
DB 61 CTGTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGCAATGGATTGGCAAGAGCCCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 181 GATTCCTGCATCAGTAGAAGACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
DB 181 GATTCCTGCATCAGTAGAAGACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
QY 241 TGAAGGATGCAGGTTATGATGATACCTCTGATTGATGACTGTGTGGATGGCTCCCAAGA 300
DB 241 TGAAGGATGCAGGTTATGATGATACCTCTGATTGATGACTGTGTGGATGGCTCCCAAGA 300
QY 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGGATGCCAGCTA 360
DB 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAATATGTTACAGCAAGNACTGAAGCTAGGAGTTTATGCAGATGTTGGAATATAA 420
DB 361 GCTAATATGTTACAGCAAGNACTGAAGCTAGGAGTTTATGCAGATGTTGGAATATAA 420
QY 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
QY 541 GCAGATGGTTATAAGCAATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
DB 541 GCAGATGGTTATAAGCAATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
```

```
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTGATGATTCCTGGAAGATATAAG 720
QY 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACAGGG 780
DB 721 AGTATCTTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTCACACTTTGGCCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTATGTCACACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 841 GTAACTCCAGATGGCCCTCTGGGCTATCATGGCTGTCTTTTATCATGTCTAATGACCTC 900
DB 841 GTAACTCCAGATGGCCCTCTGGGCTATCATGGCTGTCTTTTATCATGTCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACGAGCTTAGACAGGAGAGCAACTTTGAGGTGCG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACGAGCTTAGACAGGAGAGCAACTTTGAGGTGCG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCAGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCAGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCAGAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGAGCT 1200
DB 1141 GCCTGCTTCATCAGAGCTCTCCTGTGAAAGGAAGCTAGGTTCTATGAATGAGCT 1200
QY 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254
RESULT 8
ADJ88276
ID ADJ88276 standard; DNA; 1266 BP.
XX
AC ADJ88276;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human WT rGAL-8 (galactosidase) DNA.
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease; human; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /tag= a
FT /product= "Human WT rGAL-8 protein"
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
```


PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
DR WPI; 2004-108227/11.
DR P-PSDB; ADJ88277.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Claim 1; SEQ ID NO 11; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC DNA.
XX
SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Query Match 99.9%; Score 1252.4; DB 12; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGGCTTCCTTCCTGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGGCTTCCTTCCTGGCC 60

QY 61 CTGCTTTCTCGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCCT 120
Db 61 CTGCTTTCTGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCCT 120

QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAAGAGGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAAGAGGCCA 180

QY 181 GATTCTGCATCAGTGAAGACTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGCATCAGTGAAGACTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 241 TGAAGGATGACAGTTATGATGATACCTCTGATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGAAGGATGACAGTTATGATGATACCTCTGATGATGATGATGATGATGATGATGATGAT 300

QY 301 GATTTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360
Db 301 GATTTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360

QY 361 GCTAAATATGTTACAGCAAGAGACTGAAGTATGATGATGATGATGATGATGATGATGAT 420
Db 361 GCTAAATATGTTACAGCAAGAGACTGAAGTATGATGATGATGATGATGATGATGATGAT 420

QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTCAGTCCAGAGCTTTGCT 480
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGGACATTCAGTCCAGAGCTTTGCT 480

QY 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTGTACTGTGACAGTTGGAAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTGTACTGTGACAGTTGGAAAAATTTG 540

QY 541 GCAGATGTTTATAAGACACATGCTCTGGCCCTGAATAGGACTGCGCAGAACGATTTGTGAC 600
Db 541 GCAGATGTTTATAAGACACATGCTCTGGCCCTGAATAGGACTGCGCAGAACGATTTGTGAC 600

QY 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTCGATGATTTCTCGAAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTCGATGATTTCTCGAAAAAGTATAAG 720

QY 721 AGTATCTTGACCTGGACATCTTTTAACACAGAGAGAAATTTGTGATGTTGCTGGACACGGG 780
Db 721 AGTATCTTGACCTGGACATCTTTTAACACAGAGAGAAATTTGTGATGTTGCTGGACACGGG 780

QY 781 GGTGGAATGACCCAGATATTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960

QY 961 CAGGACCCCTTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTTGGGCAAGCAAGGGTACCAGCTTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020

QY 1021 GAACGAGCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
Db 1021 GAACGAGCTCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080

QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCCTGTAATCCT 1140

QY 1141 GCCTGCTTATCACAACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGAGCT 1200
Db 1141 GCCTGCTTATCACAACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGAGCT 1200

QY 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 9
ADM48680

ID ADM48680 standard; DNA; 1266 BP.

XX AC ADM48680;

XX AC (first entry)

DT 03-JUN-2004 (first entry)

XX Human wild type rGAL-8 DNA.

DE Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;

XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;

KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;

XX gene; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

PH 1..1266

FT CDS

FT /*tag= a

```

FT      /product= "GAL-8 protein"
XX      US2004023281-A1.
XX      05-FEB-2004.
XX      23-JUN-2003; 2003US-00602220.
XX      26-FEB-1988; 88US-00160766.
XX      26-FEB-1988; 88US-00160771.
XX      17-FEB-1989; 89US-00310881.
XX      22-OCT-1990; 90US-00600244.
XX      31-JUL-1992; 92US-00923692.
XX      30-DEC-1992; 92US-00997733.
XX      29-DEC-1993; 93US-00176414.
XX      19-JAN-1994; 94US-00184237.
XX      14-OCT-1994; 94US-00324003.
XX      21-MAY-1999; 99US-00316572.
XX      26-JUL-2000; 2000US-00628127.
XX      13-NOV-2001; 2001US-00993059.
XX      (TURP/) TURPEN T H.
XX      (KUMA/) KUMAGAI M H.
XX      (POGU/) POGUE G P.
XX      (ERWI/) ERWIN R L.
XX      (GRIL/) GRILL L K.
XX      Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX      WPI; 2004-142650/14.
XX      P-PSDB; ADM48681.
XX      New alpha-galactosidase polypeptides, useful in producing recombinant
XX      lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX      as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX      disease.
XX      Disclosure; SEQ ID NO 11; 72pp; English.
XX      The present invention relates to novel galactosidase (Gal) proteins such
XX      as rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions
XX      of the present invention are useful for producing recombinant lysosomal
XX      enzymes for enzyme replacement therapy for treating human and animal
XX      lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX      disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX      human wild type rGAL-8 DNA used in the exemplification of the invention.
XX      Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
XX      Query Match 99.9%; Score 1252.4; DB 12; Length 1266;
XX      Best Local Similarity 99.9%; Pred. No. 0;
XX      Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 ATCGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
DB      |||||||
QY      1 ATCGAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC 60
DB      |||||||
QY      61 CTGCTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGATGGCAAGACGCTT 120
DB      |||||||
QY      61 CTGCTTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGATGGCAAGACGCTT 120
DB      |||||||
QY      121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
DB      |||||||
QY      121 ACCATGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
DB      |||||||
QY      181 GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB      |||||||
QY      181 GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB      |||||||
QY      241 TGAAGAGATGCAGGTATAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAAGA 300
DB      |||||||
QY      241 TGAAGAGATGCAGGTATAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAAGA 300
DB      |||||||

```

```

QY      301 GATTTCAGAGGCAGAGCTTCAGGCAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB      |||||||
QY      301 GATTTCAGAGGCAGAGCTTCAGGCAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
DB      |||||||
QY      361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAATAAA 420
DB      |||||||
QY      361 GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAATAAA 420
DB      |||||||
QY      421 ACCTGGCAGGGCTTCCTCGGAGTTTGGATACTACGACATTCGATCCAGAGACTTGTGCT 480
DB      |||||||
QY      421 ACCTGGCAGGGCTTCCTCGGAGTTTGGATACTACGACATTCGATCCAGAGACTTGTGCT 480
DB      |||||||
QY      481 GACTGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540
DB      |||||||
QY      481 GACTGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 540
DB      |||||||
QY      541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
DB      |||||||
QY      541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
DB      |||||||
QY      601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB      |||||||
QY      601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB      |||||||
QY      661 CAGTACTGCAATCACCTGGCGAAATTTTGTCTGACATTTGATGATGATTTCTCTGGAAGTATAAAG 720
DB      |||||||
QY      661 CAGTACTGCAATCACCTGGCGAAATTTTGTCTGACATTTGATGATGATTTCTCTGGAAGTATAAAG 720
DB      |||||||
QY      721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGACACAGGG 780
DB      |||||||
QY      721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGCTGACACAGGG 780
DB      |||||||
QY      781 GGTGGAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTTGGAAATCAGCAA 840
DB      |||||||
QY      781 GGTGGAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTTGGAAATCAGCAA 840
DB      |||||||
QY      841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTAAAGACCTC 900
DB      |||||||
QY      841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTAAAGACCTC 900
DB      |||||||
QY      901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGGACGTTAATGGCCATCAAT 960
DB      |||||||
QY      901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGGACGTTAATGGCCATCAAT 960
DB      |||||||
QY      961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB      |||||||
QY      961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB      |||||||
QY      1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
DB      |||||||
QY      1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGCAGGAGATTGGT 1080
DB      |||||||
QY      1081 GGAACCTGCTCTTTATACCATCGCAGTTGCTTCCTGGGTTAAGGAGTGGCTGTAAATCCT 1140
DB      |||||||
QY      1081 GGAACCTGCTCTTTATACCATCGCAGTTGCTTCCTGGGTTAAGGAGTGGCTGTAAATCCT 1140
DB      |||||||
QY      1141 GCCTGCTTCATACACAGCTCCCTCTGCAAGGAAGCTAGGTTCTATGATGAGACT 1200
DB      |||||||
QY      1141 GCCTGCTTCATACACAGCTCCCTCTGCAAGGAAGCTAGGTTCTATGATGAGACT 1200
DB      |||||||
QY      1201 TCAAGGTTAAGAACTCACATAAAATCCCAAGGCACTGTTTTGCTTCAGCTATAA 1254
DB      |||||||
QY      1201 TCAAGGTTAAGAACTCACATAAAATCCCAAGGCACTGTTTTGCTTCAGCTATAA 1254
DB      |||||||

```

RESULT 10

ADU66915

ID ADU66915 standard; DNA; 1266 BP.

XX AC ADU66915;

XX DT 10-FEB-2005 (first entry)

XX


```
XX 11-AUG-2005 (first entry)
XX Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.
XX
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX Genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1266
XX /tag= a
XX /product= "Human alpha-galactosidase protein"
XX
XX US2005125859-A1.
XX
XX 09-JUN-2005.
XX
XX 08-NOV-2004; 2004US-00984389.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2005-404004/41.
XX
XX P-PSDB; AEA27445.
XX
XX New isolated polypeptides useful for producing lysosomal enzymes in
XX plants to be utilized in enzyme replacement therapy or for the
XX therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 11; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase DNA.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
XX
XX
XX Query Match 99.9%; Score 1252.4; DB 14; Length 1266;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTCGGCC 60
XX |||
XX 1 ATGCAGCTGAGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTCGGCC 60
XX
XX 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGGCAATGGATTGGCAAGACGCGCT 120
XX |||
XX 61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGGCAATGGATTGGCAAGACGCGCT 120
XX
XX 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
XX |||
XX 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
XX
XX 181 GATTCTGTCATCAGTAGAGAGCTCTTCATGGAGATGSCAGAGCTCATGGTCTCAGAGGC 240
```

```
181 GATTCTGTCATCAGTAGAGAGCTCTTCATGGAGATGSCAGAGCTCATGGTCTCAGAGGC 240
241 TGGAAGGATGCAAGTTATGAGTACCTTCGCAATGATGACTGTGGATGGCTCCCAAGA 300
241 TGGAAGGATGCAAGTTATGAGTACCTTCGCAATGATGACTGTGGATGGCTCCCAAGA 300
301 GATTGAGAGGAGAGCTTCAGGACAGACCTTCAGCGCTTTCCTCATGGGATTCGCAAGCTA 360
301 GATTGAGAGGAGAGCTTCAGGACAGACCTTCAGCGCTTTCCTCATGGGATTCGCAAGCTA 360
361 GCTAATTATGTTTACAGCAAGAGAGCTGAAAGCTAGGAGTTTATGCAAGATTTGGAATAAA 420
361 GCTAATTATGTTTACAGCAAGAGAGCTGAAAGCTAGGAGTTTATGCAAGATTTGGAATAAA 420
421 ACTGCGCAGGCTTCCTGGGAGTTTGGATAGTACTAGCAATTTGATGAGTGGTTCCTGCTT 480
421 ACTGCGCAGGCTTCCTGGGAGTTTGGATAGTACTAGCAATTTGATGAGTGGTTCCTGCTT 480
481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGATTTGGAATAATTG 540
481 GACTGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGATTTGGAATAATTG 540
541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAAGAGCAATTTGTATC 600
541 GCAGATGGTTATAAGCACATGTCCTTGGCCCTGAAATAGGACTGGCAAGAGCAATTTGTATC 600
601 TCTGTGAGTGGCTCTTTATATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
601 TCTGTGAGTGGCTCTTTATATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCAGTGGCGAAATTTGCTGCAATGATGATTCCTCGGAAAGTATAAAG 720
661 CAGTACTGCAATCAGTGGCGAAATTTGCTGCAATGATGATTCCTCGGAAAGTATAAAG 720
721 AGTATCTTGGAGCTGGACATCTTTTAAACAGAGAGAGAAATTTGATGATTCCTGCAAGGG 780
721 AGTATCTTGGAGCTGGACATCTTTTAAACAGAGAGAGAAATTTGATGATTCCTGCAAGGG 780
781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGCTTCAGCTGGATCAGCAA 840
781 GGTGGAATGACCCAGATATGTTAGTATGTCGCACTTTGGCTTCAGCTGGATCAGCAA 840
841 GTAACCTCAGATGCGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
841 GTAACCTCAGATGCGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACGGGAGACCACTTTGAAGTGTGG 1020
961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACGGGAGACCACTTTGAAGTGTGG 1020
1021 GAACGACCTCTCTCAGGCTTAGGCTGCGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
1021 GAACGACCTCTCTCAGGCTTAGGCTGCGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCTGGGTTAAGAGAGTGGCTGTATTCCT 1140
1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCTGGGTTAAGAGAGTGGCTGTATTCCT 1140
1141 GCCTGCTTATCAGACAGCTCTCCCTGTGAAAGAGAGTGGTTCATGATGAGTGGACT 1200
1141 GCCTGCTTATCAGACAGCTCTCCCTGTGAAAGAGAGTGGTTCATGATGAGTGGACT 1200
1201 TCAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTATATA 1254
1201 TCAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTTCAGCTATATA 1254
```


ID XX ADD84746 standard; DNA; 1278 BP.
AC XX ADD84746;
DT XX 29-JAN-2004 (first entry)
DE XX Human alpha-galactosidase rGAL-4 DNA.
KW XX Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
KW XX enzyme replacement therapy; lysosomal disease.
OS XX Homo sapiens.
FH XX Key Location/Qualifiers
FT CDS 1..1278
FT FT /*tag= a
FT FT /product= "Human rGAL-4"
XX US2003106095-A1.
XX 05-JUN-2003.
XX 20-MAR-2002; 2002US-00103327.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2003-801257/75.
XX P-PSDB; ADD84747.
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX Claim 1; SEQ ID NO 7; 77pp; English.
XX The invention relates to human alpha-galactosidase derivatives and the
XX nucleic acids encoding them. The polypeptides are used in a method for
XX producing active recombinant human and animal lysosomal enzymes in a
XX plant expression system. The enzymes can be used in enzyme replacement
XX therapy for the therapeutic treatment of human and animal lysosomal
XX diseases. This sequence represents DNA encoding a human alpha-
XX galactosidase derivative polypeptide of the invention.
XX
XX Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;
Query Match 99.9%; Score 1252.4; DB 10; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTGGCC 60
QY 61 CTGCTTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
DB 61 CTGCTTCTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGATTGGCAAGGCGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCC 180
DB 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCC 180
QY 181 GATTCTGCAATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB 181 GATTCTGCAATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGGNAGGATGCAGGTATAGTACCTCTGCAATTGATGCTGTTGGATGGCTCCCCAAAGA 300

RESULT 14
ADJ88272
ID ADJ88272 standard; DNA; 1278 BP.
XX
AC ADJ88272;

DB 241 TGGNAGGATGCAGGTATAGTACCTCTGCAATTGATGCTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGAGGAGAGCTTCAGGACAGCCCTCAGCGCTTCCCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAAGAGGAGAGCTTCAGGACAGCCCTCAGCGCTTCCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTTAATTATGTTTACAGCAAGAGGAGCTGAAGCTAGGAGTTATGACAGATGTGGAAATAA 420
DB 361 GCTTAATTATGTTTACAGCAAGAGGAGCTGAAGCTAGGAGTTATGACAGATGTGGAAATAA 420
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTCATGATGATGCTTGTGCT 480
DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTCATGATGATGCTTGTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGATTTGGAAATTTG 540
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGATTTGGAAATTTG 540
QY 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTGTAC 600
DB 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAGAGCATTTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATGATGATGATGATGATGAT 720
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATGATGATGATGATGATGAT 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGATGATGCTGACACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGATGATGCTGACACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGGCTCAGCTGGAATCAGCAA 840
DB 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAACTTTGGCTCAGCTGGAATCAGCAA 840
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
QY 901 CGACATCATGCGCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAATCAAT 960
DB 901 CGACATCATGCGCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTTTATACCATCGCACTTCCCTCGGTTAAAGAGTGGCTGTAAATCCT 1140
DB 1081 GGAACCTCGCTTTATACCATCGCACTTCCCTCGGTTAAAGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGTCTTCATCAACAGCTCTCCTCTGTGAAAGGAAAGTAGGGTTCTATGATGAGACT 1200
DB 1141 GCCTGTCTTCATCAACAGCTCTCCTCTGTGAAAGGAAAGTAGGGTTCTATGATGAGACT 1200
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTGTCTCAGCTATAA 1254
DB 1201 TCAAGGTTAAGAGTACATAAATCCACAGGACCTGTTTGTCTCAGCTATAA 1254

[illegible]

Db 1141 GCTGCTTCATCACAGCTCCTCCCTGTGAAGGAGCTAGGGTTCTATGAATGCACT 1200
QY 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTAGAA 1254

RESULT 15

ADM48676
ID ADM48676 standard; DNA; 1278 BP.
XX AC ADM48676;
XX 03-JUN-2004 (first entry)
DT DE Human wild type rGAL-4 DNA.
XX KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
XX gene; ds.
XX OS Homo sapiens.

PH Key Location/Qualifiers
FT CDS 4..1278
FT /*tag= a
FT /product= "GAL-4 protein"
FT /partial
FT /note= "No start codon"

US2004023281-A1.

XX PD 05-FEB-2004.
XX PF 23-JUN-2003; 2003US-00602220.
XX PR 26-FEB-1988; 88US-00160766.
XX PR 26-FEB-1988; 88US-00160771.
XX PR 17-FEB-1989; 89US-00310881.
XX PR 22-OCT-1990; 90US-00600244.
XX PR 31-JUL-1992; 92US-00923692.
XX PR 30-DEC-1992; 92US-00997733.
XX PR 28-DEC-1993; 93US-00176414.
XX PR 19-JAN-1994; 94US-00184237.
XX PR 14-OCT-1994; 94US-00324003.
XX PR 21-MAY-1999; 99US-00316572.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.

XX (TURP/) TURPEN T H.
XX PA (KUMA/) KUMAGAI M H.
XX PA (POGU/) POGUE G P.
XX PA (ERWI/) ERWIN R L.
XX PA (GRIL/) GRILL L K.

XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48677.

XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.

XX Disclosure; SEQ ID NO 7; 72pp; English.

XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal

CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-4 DNA used in the exemplification of the invention.

XX Sequence 1278 BP; 325 A; 293 C; 333 G; 327 T; 0 U; 0 Other;

QY Query Match 99.9%; Score 1252.4; DB 12; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCCCTGGCC 60

Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCCCTGGCC 60

QY 61 CTGCTTTCTCTGGGACATCCCTGGGCTAGAGCACTGCAATGATTTGGCAAGGACGCTT 120

Db 61 CTGCTTTCTCTGGGACATCCCTGGGCTAGAGCACTGCAATGATTTGGCAAGGACGCTT 120

QY 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGCACGGGAAGAGCCA 180

Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGCACGGGAAGAGCCA 180

QY 181 GATTCCTGCATCAGTGAGAAAGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240

Db 181 GATTCCTGCATCAGTGAGAAAGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240

QY 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATGATGACTGTTGGATGGCTCCCAAGA 300

Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATGATGACTGTTGGATGGCTCCCAAGA 300

QY 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTCGCCAGCTA 360

Db 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTCGCCAGCTA 360

QY 361 GCTAATTTATGTTCAAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAAATAA 420

Db 361 GCTAATTTATGTTCAAGCAAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAAATAA 420

QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGAGCTTTGCT 480

Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGAGCTTTGCT 480

QY 481 GACTGGGGAGTAGACTCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAAATTG 540

Db 481 GACTGGGGAGTAGACTCTGCTAAAATTTGATGTTGTTACTGTGACAGTTTGGAAAAATTG 540

QY 541 GCAGATGTTTAAAGCACATGCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600

Db 541 GCAGATGTTTAAAGCACATGCTTGGCCCTCAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

Db 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

QY 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720

Db 661 CAGTACTGCAATCACTGGCGAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAG 720

QY 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAAATTTGATGTTGCTGGACCCAGGG 780

Db 721 AGTATCTTGACCTGGACATCTTTTAAACCAGGAGAAATTTGATGTTGCTGGACCCAGGG 780

QY 781 GGTGGAAATCACCCAGATATGTTAGTGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840

Db 781 GGTGGAAATCACCCAGATATGTTAGTGGAACTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 841 GTAACTCAGATGGCCCTCTCGGCTTATCATGGCTGCTCTTTTATTTCTTAATGACCTC 900

Db 841 GTAACTCAGATGGCCCTCTCGGCTTATCATGGCTGCTCTTTTATTTCTTAATGACCTC 900

QY 901 CGACATCAGCCCTCAGGCAAGCTCTCTTTCAGGTAAGGACGTAATTTGCCATCAAT 960

Db 901 CGACATCAGCCCTCAGGCAAGCTCTCTTTCAGGTAAGGACGTAATTTGCCATCAAT 960

```
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1020
Db |||||
Qy 1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACGAGATTTGGT 1080
Db |||||
Qy 1021 GAACGACCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACGAGATTTGGT 1080
Db |||||
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTATTCCT 1140
Db |||||
Qy 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTATTCCT 1140
Db |||||
Qy 1141 GCCTGCTTCAACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1141 GCCTGCTTCAACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db |||||
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254
Db |||||
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACCTGTTTGTCTCAGCTATAA 1254
Db |||||
```

Search completed: December 31, 2005, 17:36:17
Job time : 848 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 15:58:11 ; Search time 5614 Seconds
(without alignments)
10450.837 Million cell updates/sec

Title: US-10-602-220-15

Perfect score: 1254
Sequence: 1 atgcagctgaggaccacaga.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252.4	99.9	1266	4	CR607242 full-leng
2	1252.4	99.9	1277	4	CR605654 full-leng
3	1248	99.5	1253	4	CR617861 full-leng
4	1160.4	92.5	1290	10	AY408540 full-leng
5	1140	90.9	1230	10	AY408541 Pan trogl
6	968.6	77.2	1086	1	AL554978 AL554978
7	926.8	73.9	1014	5	BX354096 BX354096
8	922.8	73.6	999	1	AL52630 AL52630
9	915.8	73.0	1067	3	BM564282 AGENCOURT
10	915	73.0	1026	1	AL577581 AL577581
11	897.2	71.5	1376	4	AK054547 Mus muscu
12	897.2	71.5	2962	4	AK040081 Mus muscu
13	883.2	70.4	922	7	CO645623 ILLUMIGEN
14	881.8	70.3	1133	7	CO645672 ILLUMIGEN
15	874.8	69.8	1071	1	AL575861 AL575861
16	874.6	69.7	1005	3	BQ062192 AGENCOURT
17	870	69.4	984	7	CO646251 ILLUMIGEN
18	862.4	68.8	920	7	CO645464 ILLUMIGEN
19	856.4	68.3	923	7	CT005156 CT005156
20	826	65.9	1236	10	AY408542 Mus muscu
21	814.2	64.9	976	5	BQ956043 AGENCOURT
22	807.8	64.4	937	5	BQ934640 AGENCOURT

23	797	63.6	922	5	BUS40848	BUS40848 AGENCOURT
24	794.8	63.4	898	5	BUI91867	BUI91867 AGENCOURT
25	783.8	62.5	888	5	BUI54569	BUI54569 AGENCOURT
26	782	62.4	852	6	CA454083	CA454083 AGENCOURT
27	768	61.2	801	5	BUS96617	BUS96617 AGENCOURT
28	765.4	61.0	927	2	BE622583	BE622583 601440703
29	765.2	61.0	813	5	BX344841	BX344841 BX344841
30	761.4	60.7	871	6	CA454143	CA454143 AGENCOURT
31	757.6	60.4	1025	3	BM450649	BM450649 AGENCOURT
32	752.2	60.0	958	6	CA487415	CA487415 AGENCOURT
33	747.6	59.6	849	6	CA487531	CA487531 AGENCOURT
34	747	59.6	796	8	DR156332	DR156332 HESC2_64
35	747	59.6	845	7	CR983967	CR983967 CR983967
36	746.8	59.6	908	2	BE379425	BE379425 601237275
37	742.6	59.2	782	2	BI224248	BI224248 602940538
38	742	59.2	871	3	BQ225444	BQ225444 AGENCOURT
39	741.4	59.1	1066	2	BG824323	BG824323 602727346
40	741.2	59.1	921	6	CA488935	CA488935 AGENCOURT
41	737.2	58.8	879	2	BG824387	BG824387 602728627
42	729	58.1	749	5	BX344842	BX344842 BX344842
43	727.6	58.0	857	5	BX374627	BX374627 BX374627
44	704.4	56.2	800	6	CB529027	CB529027 UI-H-FT2-
45	698.8	55.7	767	3	BI753664	BI753664 603023559

ALIGNMENTS

RESULT 1
CR607242

LOCUS
DEFINITION

1266 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CS0DI067YJ01 of Placenta Cot 25-normalized
of Homo sapiens (human).

ACCESSION
CR607242

VERSION
HTC; CDSLT.CDNA.

KEYWORDS
Homo sapiens (human)

SOURCE
Homo sapiens

ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 1266)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1266)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES

source

Location/Qualifiers
1..1266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI067YJ01"
/issue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 99.9%; Score 1252.4; DB 4; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 60
Db 7 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 66
QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
Db 67 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 126
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 180
Db 127 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 186
QY 181 GATTCCTGTCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 187 GATTCCTGTCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 246
QY 241 TGAAGGATGCAGGTTTATGAGTACCTCTGCAATTGATGACTGTTGGATGCTCCCCAAAGA 300
Db 247 TGAAGGATGCAGGTTTATGAGTACCTCTGCAATTGATGACTGTTGGATGCTCCCCAAAGA 306
QY 301 GATTCAGAGGCGAGACTTCAGGCGAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
Db 307 GATTCAGAGGCGAGACTTCAGGCGAGACCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 366
QY 361 GCTAATTATGTTCCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 420
Db 367 GCTAATTATGTTCCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 426
QY 421 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACCTACGACATTCGATGCCAGACCTTTGCT 480
Db 427 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATACCTACGACATTCGATGCCAGACCTTTGCT 486
QY 481 GACTGGGGAGTAGACTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Db 487 GACTGGGGAGTAGACTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 546
QY 541 GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGCGCAGAGCATTTGTGTAC 600
Db 547 GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGCGCAGAGCATTTGTGTAC 606
QY 601 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATCAGAAATCCGA 660
Db 607 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATCAGAAATCCGA 666
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAGTATAAAG 720
Db 667 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTGGAAAGTATAAAG 726
QY 721 AGTATCTTGACCTGGACATCTTTTAAACCAGAGAGAAATTTGTCATGTTGTCGACACAGG 780
Db 727 AGTATCTTGACCTGGACATCTTTTAAACCAGAGAGAAATTTGTCATGTTGTCGACACAGG 786
QY 781 GGTTCGAATGACCCAGATATGTTAGTATGGCAACTTTGGCTTCAGCTGGAATCAGCAA 840
Db 787 GGTTCGAATGACCCAGATATGTTAGTATGGCAACTTTGGCTTCAGCTGGAATCAGCAA 846
QY 841 GTAACCTCAGATGGCCCTTCGGCTATCATGGCTGCTCTTTATTTCATGCTAATGACCTC 900
Db 847 GTAACCTCAGATGGCCCTTCGGCTATCATGGCTGCTCTTTATTTCATGCTAATGACCTC 906
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGCTAATTTGGCATCAAT 960
Db 907 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGCTAATTTGGCATCAAT 966
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 967 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTTAGACAGGAGACAACTTTGAAGTGTGG 1026
QY 1021 GAAAGAGCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGATTTGGT 1080
Db 1027 GAAAGAGCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGATTTGGT 1086
QY 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTCTGTAATCCT 1140
```

```
Db 1087 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGTAAAGGAGTGGCTGTAATCCT 1146
QY 1141 GCTGCTTTCATCACACAGCTTCTCCCTGTGAAAGGAAGCTAGGTTCTTATGAATGACT 1200
Db 1147 GCTGCTTTCATCACACAGCTTCTCCCTGTGAAAGGAAGCTAGGTTCTTATGAATGACT 1206
QY 1201 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254
Db 1207 TCAAGGTTAAGAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTAGAA 1260

RESULT 2
CR605654 1277 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSDB008YB11 of Neuroblastoma Cot
DEFINITION 10-normalized of Homo sapiens (human).
ACCESSION CR605654
VERSION CR605654.1 GI:50486461
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1277)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLES Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue

REFERENCE 2 (bases 1 to 1277)
AUTHORS Genoscope.
TITLES Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

FEATURES
source
1..1277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDB008YB11"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid_id="pCMVSPORT_6"

ORIGIN
Query Match 99.9%; Score 1252.4; DB 4; Length 1277;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 60
Db 10 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGCTTCCCTGGCC 69
QY 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 120
Db 70 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCT 129
QY 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 180
Db 130 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 189
QY 181 GATTCCTGTCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 190 GATTCCTGTCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 249
```

```
Qy 241 TGAAGAGTGCAGGTTATGATGCTCTGCTGATGATGCTGTTGGATGGCTCCCAAGA 300
Db 250 TGAAGAGTGCAGGTTATGATGCTCTGCTGATGATGCTGTTGGATGGCTCCCAAGA 309
Qy 301 GATTTCAGAGGCGACATTCAGGAGACCTCAGCGCTTCTCATGGATTTCGCGAGCTA 360
Db 310 GATTTCAGAGGCGACATTCAGGAGACCTCAGCGCTTCTCATGGATTTCGCGAGCTA 369
Qy 361 GCTAATATGTTTACAGCAAGGCTGAAGCTAGGAGTTTATGCGATGTTTGGAAATAA 420
Db 370 GCTAATATGTTTACAGCAAGGCTGAAGCTAGGAGTTTATGCGATGTTTGGAAATAA 429
Qy 421 ACCTGCGAGCTTCCCTGGAGTTTGGATACACAGATTGATGATGATGATGATGATGAT 480
Db 430 ACCTGCGAGCTTCCCTGGAGTTTGGATACACAGATTGATGATGATGATGATGATGAT 489
Qy 481 GACTGGGAGTGTAGTCTGCTAAATTTGATGTTTGTGTTTGTGACAGCTTTGGAAATTTG 540
Db 490 GACTGGGAGTGTAGTCTGCTAAATTTGATGTTTGTGTTTGTGACAGCTTTGGAAATTTG 549
Qy 541 GCAGATGGTTTATAGCACATGCTCTTGGCCCTGAAATAGGACTGCGAGAGCATTTGTGTAC 600
Db 550 GCAGATGGTTTATAGCACATGCTCTTGGCCCTGAAATAGGACTGCGAGAGCATTTGTGTAC 609
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTCAAAGGCCAATATATACAGAAATCCGA 660
Db 610 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTCAAAGGCCAATATATACAGAAATCCGA 669
Qy 661 CAGTACTGCAATCACTGCGGAAATTTGCTGACATGATGATGATGATGATGATGATGATGAT 720
Db 670 CAGTACTGCAATCACTGCGGAAATTTGCTGACATGATGATGATGATGATGATGATGATGAT 729
Qy 721 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATGTTGATGTTTGTGACACAGGG 780
Db 730 AGTATCTTGGACTGACATCTTTTAAACAGGAGAGAAATGTTGATGTTTGTGACACAGGG 789
Qy 781 GTTTGGAAATGACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 790 GTTTGGAAATGACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGAT 849
Qy 841 GTRACTCAGATGGCCCTCGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 900
Db 850 GTRACTCAGATGGCCCTCGGGCTATCATGCTGCTCTCTTTTATTCATGCTTAATGACCTC 909
Qy 901 CGACATCATGACCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGCTTAATGGCATCAAT 960
Db 910 CGACATCATGACCCCTCAAGCAAGCTCTCTTCCAGGATAAGGACGCTTAATGGCATCAAT 969
Qy 961 CAGAACCCCTTGGGCAAGGATACAGCTTACAGGAGAGCAAACTTTGAAGTGTGG 1020
Db 970 CAGAACCCCTTGGGCAAGGATACAGCTTACAGGAGAGCAAACTTTGAAGTGTGG 1029
Qy 1021 GAACGACCTCTCAGGCTTGTAGCTGGCTGTAGCTATGATTAACCGGAGGAGATTGGT 1080
Db 1030 GAACGACCTCTCAGGCTTGTAGCTGGCTGTAGCTATGATTAACCGGAGGAGATTGGT 1089
Qy 1081 GGACCTCGCTCTTATACATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1090 GGACCTCGCTCTTATACATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1149
Qy 1141 GCCTGCTTATCAACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTCTATGATGAGCT 1200
Db 1150 GCCTGCTTATCAACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTCTATGATGAGCT 1209
Qy 1201 TCAAGTTTAAAGTACATATAATCCACAGGCACTGTTTGTCTGAGCTATA 1254
Db 1210 TCAAGTTTAAAGTACATATAATCCACAGGCACTGTTTGTCTGAGCTATA 1263
```

RESULT 3
CR617861 1253 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0DK009YE12 of HeLa cells Cot 25-normalized
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

of Homo sapiens (human).

CR617861

CR617861.1 GI:50498668

HTC; CNSLT_cDNA.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1253)

Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact : feng liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/invitrogen Corporation 1600

Paradise Avenue

2 (bases 1 to 1253)

Genoscope.

Direct Submission

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.

Location/Qualifiers

source

1..1253

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK009YE12"

/issue_type="HeLa cells Cot 25-normalized"

/plasmid="pCMVSPORT_6"

ORIGIN

Query Match 99.5%; Score 1248; DB 4; Length 1253;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCCTCGGCC 60
Db 6 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCCCTCGGCC 65
Qy 61 CTGCTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
Db 66 CTGCTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 125
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 126 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 185
Qy 181 GATTCTGCTCATCTGAGTACCTCTCATGAGATGCGAGAGCTCATGTCTCAGAGGC 240
Db 186 GATTCTGCTCATCTGAGTACCTCTCATGAGATGCGAGAGCTCATGTCTCAGAGGC 245
Qy 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCTGATGATGATGATGATGATGATGATGATGAT 300
Db 246 TGAAGAGTGCAGGTTATGAGTACCTCTGCTGATGATGATGATGATGATGATGATGATGAT 305
Qy 301 GATTTCAGAGGCGACATTCAGGAGACCTCAGCGCTTTCCTCATGCGATTTCGCGAGCTA 360
Db 306 GATTTCAGAGGCGACATTCAGGAGACCTCAGCGCTTTCCTCATGCGATTTCGCGAGCTA 365
Qy 361 GCTAATATGTTTACAGCAAGGCTGAAGCTAGGAGTTTATGCGATGTTTGGAAATAA 420
Db 366 GCTAATATGTTTACAGCAAGGCTGAAGCTAGGAGTTTATGCGATGTTTGGAAATAA 425
Qy 421 ACCTGCGAGGCTTCCCTGGAGTTTGGATACACAGATTGATGATGATGATGATGATGATGAT 480
Db 426 ACCTGCGAGGCTTCCCTGGAGTTTGGATACACAGATTGATGATGATGATGATGATGATGAT 485
```

QY 481 GACTGGGGAGTACATCTGCTAAATTTGATGTTGTACTGTGACAGTTGGGAAAATTTG 540
Db 486 GACTGGGGAGTACATCTGCTAAATTTGATGTTGTACTGTGACAGTTTGGGAAAATTTG 545
QY 541 GCAGATGGTTATAGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Db 546 GCAGATGGTTATAGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 605
QY 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 606 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 665
QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720
Db 666 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 725
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 726 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 785
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTGGCCCTCAGCTGGATCAGCAA 840
Db 786 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTGGCCCTCAGCTGGATCAGCAA 845
QY 841 GTAACCTCAGATGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTAATGACCTC 900
Db 846 GTAACCTCAGATGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTAATGACCTC 905
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGCAAGCTTAATTCATCAAT 960
Db 906 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGCAAGCTTAATTCATCAAT 965
QY 961 CAGAACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 966 CAGAACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1025
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAACCGCAGGAGATGGT 1080
Db 1026 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAACCGCAGGAGATGGT 1085
QY 1081 GGACCTCGCTCTTATACATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1140
Db 1086 GGACCTCGCTCTTATACATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTCTTAATCCT 1145
QY 1141 GCCTGCTTCATCAGACGCTCCCTCTGTGAAAGGAGTGGTTCATGATGAGT 1200
Db 1146 GCCTGCTTCATCAGACGCTCCCTCTGTGAAAGGAGTGGTTCATGATGAGT 1205
QY 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAG 1248
Db 1206 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAG 1253

RESULT 4
LOCUS AY408540 1290 bp DNA linear GSS 15-DEC-2003
DEFINITION Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY408540
VERSION AY408540.1 GI:39764511
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 1290)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source 1..1290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xrefs="taxon:9606"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"

ORIGIN
Query Match 92.5%; Score 1160.4; DB 10; Length 1290;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCCTGCATCAGTCAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
QY 241 TGGAGGATGTCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGAGTGGCTCCCAAGA 300
Db 241 TGGAGGATGTCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGAGTGGCTCCCAAGA 300
QY 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
QY 361 GCTAATTTATGTTTACAGCAAGGAGCTGAGCTAGGAGTTTATGCAGATGTTGAAAATAA 420
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAGCTAGGAGTTTATGCAGATGTTGAAAATAA 420
QY 421 ACCTGGCGAGGCTTCCCTGGAGTTCCTTGGATGACTTACGACATGATGATGATGATGATG 480
Db 421 ACCTGGCGAGGCTTCCCTGGAGTTCCTTGGATGACTTACGACATGATGATGATGATGATG 480
QY 481 GACTGGGAGTGTAGTCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Db 481 GACTGGGAGTGTAGTCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
QY 541 GCAGATGTTTATAGGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Db 541 GCAGATGTTTATAGGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720

JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civallo,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source 1..1290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xrefs="taxon:9606"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"

ORIGIN
Query Match 92.5%; Score 1160.4; DB 10; Length 1290;
Best Local Similarity 92.6%; Pred. No. 0;
Matches 1161; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTGGCC 60
QY 61 CTGCTTTCCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 181 GATTCCTGCATCAGTCAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCCTGCATCAGTCAGAGAGCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
QY 241 TGGAGGATGTCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGAGTGGCTCCCAAGA 300
Db 241 TGGAGGATGTCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGAGTGGCTCCCAAGA 300
QY 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
QY 361 GCTAATTTATGTTTACAGCAAGGAGCTGAGCTAGGAGTTTATGCAGATGTTGAAAATAA 420
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAGCTAGGAGTTTATGCAGATGTTGAAAATAA 420
QY 421 ACCTGGCGAGGCTTCCCTGGAGTTCCTTGGATGACTTACGACATGATGATGATGATGATG 480
Db 421 ACCTGGCGAGGCTTCCCTGGAGTTCCTTGGATGACTTACGACATGATGATGATGATGATG 480
QY 481 GACTGGGAGTGTAGTCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Db 481 GACTGGGAGTGTAGTCTGCTAAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
QY 541 GCAGATGTTTATAGGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Db 541 GCAGATGTTTATAGGACACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720

```
Qy 721 AGTATCTTGAGCTGAGCATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGACACAGG 780
Db 721 AGTATCTTGAGCTGAGCATCTTTTAAACAGAGAGAAATGTTGATGTTGCTGACACAGG 780
Qy 781 GGTGGGAATGACCCAGATATGTTAGTATGGAACCTTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGGAATGACCCAGATATGTTAGTATGGAACCTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACATCAGATGGCCCTCTGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATGAAGCACTTAATGGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATGAAGCACTTAATGGCATCAAT 960
Qy 961 CAGGACCCCTTGGCAGCAAGGTCACGCTTAGCAGGAGACCACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGCAGCAAGGTCACGCTTAGCAGGAGACCACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCAGACGCTCTCCCTGTAAGAGGAGCTAGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTCATCAGACGCTCTCCCTGTAAGAGGAGCTAGGTTCTATGATGAGCT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254

RESULT 5
AY408541 1290 bp DNA linear GSS 15-DEC-2003
LOCUS Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408541
VERSION AY408541.1 GI:39764512
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1290
/organism="Pan troglodytes"
```

```
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/genes="GLA"
/locus_tag="HCM3258"

ORIGIN
Query Match 90.9%; Score 1140; DB 10; Length 1290;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1143; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTCGGCC 60
Db 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTCGGCC 60
Qy 61 CTGCTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTCTTTTCTTGGGACATCCCTTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 181 GATTCTTCGATCAGTCAGAGAGCTTTCATGGAGATGCGAGAGCTCATGCTCTCAGAGGC 240
Db 181 GATTCTTCGATCAGTCAGAGAGCTTTCATGGAGATGCGAGAGCTCATGCTCTCAGAGGC 240
Qy 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCAAGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTCAAGAGGAGAGCTTCAGGAGACCTTCAGGCTTTCTCATNGGATTCGCCAGCTA 360
Db 301 GATTCAAGAGGAGAGCTTCAGGAGACCTTCAGGAGACCTTCAGGCTTTCTCATNGGATTC 360
Qy 361 GCTAATTATGTTTCCAGAGCAAGGAGCTGAACTAGGAGATTTATGAGATGTTGGAAATAA 420
Db 361 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 420
Qy 421 ACCTGCGCAGGCTTCCTCGGAGCTTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCTCGGAGCTTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGTCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGTCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC 600
Qy 601 TCCTGTGAGTGGGCTCTTTTATATATGTCGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 661 CAGTACTGCAATCAGCTGGGAAATTTTGTCTGCAATTCATGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGGAAATTTTGTCTGCAATTCATGATGATTCCTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGGAGAGAAATGTTGATGTTGCTGGACAGGG 780
Qy 781 GGTGGGAATGACCCAGATATGTTAGTATGGAACCTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGGAATGACCCAGATATGTTAGTATGGAACCTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACATCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATGAAGCACTTAATGGCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAGGATGAAGCACTTAATGGCATCAAT 960
```

```
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCCAGGATAAGGACGTAATTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGACTTGTCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCTGCTTCATCACAGCTCTCTCCCTGTGTAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Db 1141 GCTGCTTCATCACAGCTCTCTCCCTGTGTAAGGAAGCTAGGGTCTTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATATA 1254
```

RESULT 6

```
AL554978
LOCUS AL554978 Homo sapiens HELA CELLS cot 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DK009YE12 5-PRIME, mRNA sequence.
ACCESSION AL554978
VERSION AL554978.3 GI:45859723
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE

1. (bases 1 to 1086)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31276787.

Contact: Genoscope

Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06QP1&c=10506.r.

FEATURES

source
1. .1086
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 77.2%; Score 968.6; DB 1; Length 1086;
Best Local Similarity 98.0%; Pred. No. 7.7e-279;
Matches 1012; Conservative 9; Mismatches 8; Indels 4; Gaps 4;
Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTCGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60

```
Db 6 ATGCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTGC - CTTGGCTCTCTGGCC 64
Qy 61 CTCGTTTCTTGGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGAGCGCT 120
Db 65 CTCGTTTCTTGGGACATCTCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGAGCGCT 124
Qy 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGCACCTGCCAGGAGAGCCA 180
Db 125 ACCATGGGCTGGCTGCACCTGGGAGCGCTTTCATGTGCAACCTTGCACCTGCCAGGAGAGCCA 184
Qy 181 GATTTCCTGCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGGC 240
Db 185 GATTTCCTGCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGGC 244
Qy 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATGATGACTGTGGATGGCTCCCAAGA 300
Db 245 TGGAGGATGCAGGTTATGAGTACCTCTGCATGATGACTGTGGATGGCTCCCAAGA 304
Qy 301 GATTTCAGAAAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 305 GATTTCAGAAAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 364
Qy 361 GCTAATTATGTTTCCACAGCAAGGAGCTGAAGCTAGGATTTATGACAGATGTTGGAATAAAA 420
Db 365 GCTAATTATGTTTCCACAGCAAGGAGCTGAAGCTAGGATTTATGACAGATGTTGGAATAAAA 424
Qy 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATCTAGCAATTCATGCCAGAGCTTTGCT 480
Db 425 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATCTAGCAATTCATGCCAGAGCTTTGCT 484
Qy 481 GACTGGGGAGTAGACTCTGCTTAAATTTGATGGTGTGTTACTGTGACAGTGTGGAAAAATTG 540
Db 485 GACTGGGGAGTAGACTCTGCTTAAATTTGATGGTGTGTTACTGTGACAGTGTGGAAAAATTG 544
Qy 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGGAGAGCAATGTGTAC 600
Db 545 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGGAGAGCAATGTGTAC 604
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 605 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTCAAAAGCCCAATTATACAGAAATCCGA 664
Qy 661 CAGTACTGCAATCACTGGGAAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAAG 720
Db 665 CAGTACTGCAATCACTGGGAAAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAAAG 724
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTGACACAGGG 780
Db 725 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTGACACAGGG 784
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 785 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 844
Qy 841 GTAACTCAGATGGCCCTCTGGGCTA - TCATGGCTGCTCTCTTATTTATGATGCTAATGACCT 899
Db 845 GTAACTCAGATGGCCCTCTGGGCTA - TCATGGCTGCTCTCTTATTTATGATGCTAATGACCT 904
Qy 900 CCGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAA 959
Db 905 CCGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAA 964
Qy 960 TCAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGCAGGAGAGCAACCTTTGAAGTGTG 1019
Db 965 TCAGGA - CCCYTKGGCAAGCAAGGGTA - CAGCTTAGAAGGGGGRGACAMTTTGWAGTGTG 1022
Qy 1020 GGAACGACCTCTC 1032
Db 1023 GGAGAGCTCTCTC 1035
```

RESULT 7


```

BX354096      1014 bp      mRNA      linear      EST 23-APR-2004
LOCUS          BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION     cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION      BX354096
VERSION        BX354096.2 GI:46550107
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homiidae; Homo.
REFERENCE      1 (bases 1 to 1014)
AUTHORS        Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE          Full-length cDNA libraries and normalization
JOURNAL        Unpublished (2001)
COMMENT        On May 5, 2003 this sequence version replaced gi:30379845.
               Contact: Genoscope
               Genoscope - Centre National de Sequencage
               2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
               Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
               1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
               end enriched, double-strand cDNA was digested with Not I and cloned
               into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
               was normalized. Library was constructed by Life Technologies, a
               division of Invitrogen. This sequence belongs to sequence cluster
               10506.r
               For more information about this cluster, see
               http://www.genoscope.cns.fr/cdna?s=CS0DC014DD02QP1&c=10506.r.

FEATURES
source
   1..1014
   /organism="Homo sapiens"
   /mol_type="mRNA"
   /db_xref="taxon:9606"
   /clone="CS0DC014YH04"
   /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
   /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
   /notes="1st strand cDNA was primed with a NotI-oligo(dT)
   primer. Five prime end enriched, double-strand cDNA was
   digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
   was normalized."

ORIGIN
Query Match      73.9%; Score 926.8; DB 5; Length 1014;
Best Local Similarity 98.4%; Pred. No. 2.7e-266;
Matches 958; Conservative 8; Mismatches 5; Indels 3; Gaps 3;

QY 11 GGAACCCAGAACTACATCTGGGCTGGCGCTTGGCGTTGGCTTCGCTTCCTGGCCCTCGTTTCCT 70
DB 29 GGAACCCAGAACTACATCTGGGCTGGCGCTTGGC-STTCGCTTCCTGGCCCTCGTTTCCT 87

QY 71 GGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCCTACCATGGGCT 130
DB 88 GGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGAGCGCCTACCATGGGCT 147

QY 131 GCGTCGACATGGGAGCGCTTCATGTGCAACCTTGAATGCCAGGAAGAGCCAGATTCCTGCA 190
DB 148 GCGTCGACATGGGAGCGCTTCATGTGCAACCTTGAATGCCAGGAAGAGCCAGATTCCTGCA 207

QY 191 TCAGTGAGAGCTCTTCATGAGATGGCAGACTCATGCTCTCAGAGGCTGGAAGATG 250
DB 208 TCAGTGAGAGCTCTTCATGAGATGGCAGACTCATGCTCTCAGAGGCTGGAAGATG 267

QY 251 CAGGTTATGATGATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 310
DB 268 CAGGTTATGATGATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 327

QY 311 GCAGATTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTAGCTAAATTATG 370
DB 328 GCAGATTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTAGCTAAATTATG 387

QY 371 TTCACGAAAGAGCTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 430

```

```

DB 388 TTCACGAAAGAGCTAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 447
QY 431 GCTTCCCTGGGAGTCTTTGGTACTACGACATTTGATGCCAGACCTTTTGTGCTGACTGGGAG 490
DB 448 GCTTCCCTGGGAGTCTTTGGTACTACGACATTTGATGCCAGACCTTTTGTGCTGACTGGGAG 507
QY 491 TAGATCTGTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTGGCAGATGGT- 549
DB 508 TAGATCTGTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTGGCAGATGGTGA 567
QY 550 TATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTTGTGTAATCTCTCTGTGAG 609
DB 568 TATAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTTGTGTAATCTCTCTGTGAG 627
QY 610 TGGCCTCTTTATATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGACAGTACTGC 669
DB 628 TGGCCTCTTTATATGATGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGACAGTACTGC 687
QY 670 AATCAGCTGGCGAAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 729
DB 688 AATCAGCTGGCGAAATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 747
QY 730 GACTGGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTCTGACACAGGGGTTGGAAT 789
DB 748 GACTGGACATCTTTTAAACAGAGAGAGAAATTTGTTGATGTTCTGACACAGGGGTTGGAAT 807
QY 790 GACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 849
DB 808 GACCCAGATATGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867
QY 850 ATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTATGTTATGTTATGTTATGTTATGTTATG 909
DB 868 ATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATTTATGTTATGTTATGTTATGTTATGTTATG 927
QY 910 AGCCCTCAAGCCAAAGCTCTCTCTTCAAGATAAGAGAGCTTAATGTCATCAATCAGGACCCC 969
DB 928 AGCCCTCAAGCCAAAGCTCTCTCTTCAAGATAAGAGAGCTTAATGTCATCAATCAGGACCCC 986
QY 970 TTGGGCAAGCAAGG 983
DB 987 TKGGMAGCAAGGK 1000

```

```

RESULT 8
AL552630
LOCUS
DEFINITION  AL552630 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1067XJ01 5-PRIME, mRNA sequence.
ACCESSION  AL552630
VERSION    AL552630.3 GI:45857411
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
           Homiidae; Homo.
REFERENCE   1 (bases 1 to 999)
AUTHORS     Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:31274445.
           Contact: Genoscope
           Genoscope - Centre National de Sequencage
           2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
           end enriched, double-strand cDNA was digested with Not I and cloned
           into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
           was normalized. Library was constructed by Life Technologies, a
           division of Invitrogen. This sequence belongs to sequence cluster
           10506.r
           For more information about this cluster, see
           http://www.genoscope.cns.fr/cdna?s=CS0D1067XJ01&c=10506.r.

```


FEATURES

source Location/Qualifiers
 1. .999
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS01067YJ01"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 73.6%; Score 922.8; DB 1; Length 999;
 Best Local Similarity 99.2%; Pred. No. 4.3e-265;
 Matches 976; Conservative 3; Mismatches 0; Indels 5; Gaps 5;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCGCTTCCTGGCC 60
 DB |||||
 QY 7 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGG-TCGCTTCCTGGCC 65
 DB |||||

QY 61 CTCGTTTCCTGGGACATCTCTGGGCTAGAGCACTGGCAATGATGGCAAGAGCGCT 120
 DB |||||
 QY 66 CTGTTTCCTGGGACATCTCTGGGCTAGAGCACTGGCAATGATGGCAAGAGCGCT 125
 DB |||||

QY 121 ACCATGGGCTGGTGCATCTGGGAGCGCTTCATGTGCNACCTTGACTGCCAGGAGGCCA 180
 DB |||||
 QY 126 ACCATGGGCTGGTGCATCTGGGAGCGCTTCATGTGCNACCTTGACTGCCAGGAGGCCA 185
 DB |||||

QY 181 GATTTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 DB |||||

QY 241 TGAAGGATCAGGTTATAGTACCTCTGCATTCATGATGACTGTGGATGGCTCCCAAGA 300
 DB |||||

QY 246 TGAAGGATCAGGTTATAGTACCTCTGCATTCATGATGACTGTGGATGGCTCCCAAGA 305
 DB |||||

QY 301 GATTTCAGAGGCGAGCTTCAGGAGAGCCCTCAGGCTTCCTCATGGATTCGCCAGCTA 360
 DB |||||

QY 306 GATTTCAGAGGCGAGCTTCAGGAGAGCCCTCAGGCTTCCTCATGGATTCGCCAGCTA 365
 DB |||||

QY 361 GCTAATATGTTTCACAGCAAGGAGCTCAAGCTAGGAGTTTATGCAGATGTTGGAATATA 420
 DB |||||

QY 366 GCTAATATGTTTCACAGCAAGGAGCTCAAGCTAGGAGTTTATGCAGATGTTGGAATATA 425
 DB |||||

QY 421 ACCTGGCAGGCTTCCTCGGAGTTTGGGATACATACAGATGATGATGCCAGACCTTTGCT 480
 DB |||||

QY 426 ACCTGGCAGGCTTCCTCGGAGTTTGGGATACATACAGATGATGATGCCAGACCTTTGCT 485
 DB |||||

QY 481 GACTGGGGAGTAGTCTGCTAATATTTGATGTTTGTACTGTGACAGTTTGGAAATTTG 540
 DB |||||

QY 486 GACTGGGGAGTAGTCTGCTAATATTTGATGTTTGTACTGTGACAGTTTGGAAATTTG 545
 DB |||||

QY 541 GCAGATGTTTATAGCAGCATGTCCTGGCCCTGAATAGGAGTGGCAGAGCATTTGTGTAC 600
 DB |||||

QY 546 GCAGATGTTTATAGCAGCATGTCCTGGCCCTGAATAGGAGTGGCAGAGCATTTGTGTAC 605
 DB |||||

QY 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
 DB |||||

QY 606 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 665
 DB |||||

QY 661 CAGTACTGCAATCACTGGGAAATTTGCTGACATTCATGCTTCCTGGAAAGTATATAAG 720
 DB |||||

QY 666 CAGTACTGCAATCACTGGGAAATTTGCTGACATTCATGCTTCCTGGAAAGTATATAAG 725
 DB |||||

QY 721 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGTGGACCAAGG 780
 DB |||||

QY 726 AGTATCTTGGAGTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGTGGACCAAGG 785
 DB |||||

QY 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
 DB |||||

QY 786 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 845
 DB |||||

QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
 DB |||||

QY 846 GTPAACTCAGATGGCCCTCTGGGCTATCATGGCTGCT-CVTTATTCATGTCTAATGACCTC 904
 DB |||||

QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960
 DB |||||

QY 905 CGACACATCAGCCCTCAAGCC-AAAGCTCTCCTTCAGGATAAGGACGTAATTCG-CATCAAT 962
 DB |||||

QY 961 CAGGACCCCTTGGGCAAGCAAGGG 984
 DB |||||

QY 963 CAGGA-CCCTTGGGCAAGCAAGSK 985
 DB |||||

RESULT 9
 BM564282 1067 bp mRNA linear EST 20-FEB-2002
 LOCUS AGENCOURT_6560104 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5742198
 DEFINITION 5', mRNA sequence.
 ACCESSION BM564282
 VERSION BM564282.1 GI:18811955
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1067)
 NIH-MGC http://mgc.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaabs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLMI2759 row: P column: 07
 High quality sequence stop: 761.

FEATURES

source Location/Qualifiers
 1. .1067
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5742198"
 /tissue_type="medulla"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_119"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site2: EcoRV (destroyed); RNA source normal medulla from
 anonymous male age 27. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.3 kb, insert size range
 0.9-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 013. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 73.0%; Score 915.8; DB 3; Length 1067;
 Best Local Similarity 96.4%; Pred. No. 5.6e-263;
 Matches 1009; Conservative 0; Mismatches 31; Indels 7; Gaps 7;

QY 6 GCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCCCTCGT 65
 DB |||||

QY 15 GCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCCCTCGT 74
 DB |||||

QY 66 TTCTCTGGGACATCCCTGGGGCTAGGACATGGCAATGGATTGGCAGGAGCGCTTACCAT 125
 DB |||||

QY 75 TTCTCTGGGACATCCCTGGGGCTAGGACATGGCAATGGATTGGCAGGAGCGCTTACCAT 134
 DB |||||

```
Qy 126 GGGCTGGCTGACCTGGAGGCGCTTCATGTGTGCAACCTTGACCTGCAGGAGAGCCAGATTC 185
Db 135 GGGCTGGCTGACCTGGAGGCGCTTCATGTGTGCAACCTTGACCTGCAGGAGAGCCAGATTC 194
Qy 186 CTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTATGCTCTCAGAAAGCTGGAA 245
Db 195 CTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTATGCTCTCAGAAAGCTGGAA 254
Qy 246 GGATGCAAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGCTCTCCCAAGAGATTC 305
Db 255 GGATGCAAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGCTCTCCCAAGAGATTC 314
Qy 306 AGAAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGANTCGCCAGCTAGCTAA 365
Db 315 AGAAGCAGACTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGANTCGCCAGCTAGCTAA 374
Qy 366 TTATGTTTCAGCAAGCAAGGCTGAGCTAGGATTTATGCAAGTGTGGAAATTAAGCTG 425
Db 375 TTATGTTTCAGCAAGCAAGGCTGAGCTAGGATTTATGCAAGTGTGGAAATTAAGCTG 434
Qy 426 CGCAGGCTCTCCCTGGAGTTTGGATACTACGACATTTGATGCCCGCAGACCTTTTGCTGACTG 485
Db 435 CGCAGGCTCTCCCTGGAGTTTGGATACTACGACATTTGATGCCCGCAGACCTTTTGCTGACTG 494
Qy 486 GGGAGTATGCTGTCTAAATTTGATGTTGTTACTGTGACAGTGTGGAAATTTGGCAGA 545
Db 495 GGGAGTATGCTGTCTAAATTTGATGTTGTTACTGTGACAGTGTGGAAATTTGGCAGA 554
Qy 546 TGGTTATAGCAGATGCTCTTGGCCCTGATAGACTGGCAGAGCAATTTGCTACTCTG 605
Db 555 TGGTTATAGCAGATGCTCTTGGCCCTGATAGACTGGCAGAGCAATTTGCTACTCTG 614
Qy 606 TGAGTGCCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGACAGTA 665
Db 615 TGAGTGCCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGACAGTA 674
Qy 666 CTGCAATCACTGGCGAAATTTTGCTGACATTTGATTCCTTGGAAAAAGTATAAGAGTAT 725
Db 675 CTGCAATCACTGGCGAAATTTTGCTGACATTTGATTCCTTGGAAAAAGTATAAGAGTAT 734
Qy 726 CTTGGACTGCACATCTTTTAAACAGGAGAGATTTGTTGATGTTGCTGGACCA-GGGGGTT 784
Db 735 CTTGGACTGCACATCTTTTAA-CAGGAGAGATTTGTTGATGTTGCTGGACCAAGGGGGTT 793
Qy 785 GGAATGACCCAGATATGTTAGTGTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAA 844
Db 794 GGAATGACCCAGATATGTTAGTGTAGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAA 853
Qy 845 CTCAGATGGCCCTCTGGGCTATCATGCTGCT-CCTTTATTTCATGTTCTAATGACCTCCGA 903
Db 854 CTCAGATGGCCCTCTGGGCTATCATGCTGCTGCTCCTTTATTTCATGCTAATGACCTCCG 913
Qy 904 CACATCAGCCCTCAGCCAAAGCTCTCTTCAGATTAAGGACGTAATGGCATCAATCAG 963
Db 914 AACATCAACCTC-AGCCAAAGCTCTCTTTCAAGATAAGGACGT-ATTGCCATCCATCAG 971
Qy 964 GACCCCTTGGGCAAGCAAGGTA-CCAGCTTTAGACAGGAGACAACCTTTTGAAGTGTGGGA 1022
Db 972 GAACCTTGGGCAAGCAGGAGTACCCNCTTAGACCGGAAACAACTTTTAAATGTGGGA 1031
Qy 1023 ACG-ACCTCTCTCAGGCTTAGCCCTGGG 1048
Db 1032 ACGNACTCTCTCAGGCTTTAAACCGGG 1058
```

```
RESULT 10
AL577581/c
LOCUS
DEFINITION
AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED linear EST 07-APR-2004
cDNA clone CS0DK009YE12 3-PRIME, mRNA sequence.
ACCESSION
AL577581
VERSION
AL577581.3 GI:46256603
```

KEYWORDS
SOURCE
ORGANISM

EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homidae; Homo.
1 (bases 1 to 1026)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:31315840.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06NPl&c=10506.r.
Location/Qualifiers
1..1026
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YE12"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 73.0%; Score 915; DB 1; Length 1026;
Best Local Similarity 98.3%; Pred. No. 9.5e-263;
Matches 939; Conservative 4; Mismatches 10; Indels 2; Gaps 2;
Qy 289 GCTCCCCAAGAGATTCAGAGGAGCAGACTTCAGGAGACCCCTCAGCGCTTTCCTCATGGG 348
Db 955 GCTCCCCAAGAGATTCAGAA-GCAGACTTCA-GCAGACCCCTCAGCGCTTTCCTCATGGG 898
Qy 349 ATTGCGCAGCTAGCTAATATGTTTCAGCAAAAGGACTGAAGCTAGGAGTTTATGACAGT 408
Db 897 ATTGCGCAGCTAGCTAATATGTTTCAGCAAAAGGACTGAAGCTAGGAGTTTATGACAGT 838
Qy 409 GTTGGAAATAAAACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCC 468
Db 837 GTTGGAAATAAAACCTGCGCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCC 778
Qy 469 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGTTTACTGTGACAGT 528
Db 777 CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAAATTTGATGTTTACTGTGACAGT 718
Qy 529 TTGGAATAATTTGCGAGATGTTAAGCAGATCTCTTGGCCCTTGAATAGGAGTGGCAGA 588
Db 717 TTGGAATAATTTGCGAGATGTTAAGCAGATCTCTTGGCCCTTGAATAGGAGTGGCAGA 658
Qy 589 AGCATTTGTGACTCTCTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTAT 648
Db 657 AGCATTTGTGACTCTCTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTAT 598
Qy 649 ACAGAAATCCGACAGTACTGCAATCACTGCGGAAAATTTTCTGCTGACATTTGATCTCTGG 708
Db 597 ACAGAAATCCGACAGTACTGCAATCACTGCGGAAAATTTTCTGCTGACATTTGATCTCTGG 538
Qy 709 AAAAGTATTAAGAGTATCTTGGACTGACATCTTTTAAACCAGGAGAGAAATTTGTTGATGTT 768
Db 537 AAAAGTATTAAGAGTATCTTGGACTGACATCTTTTAAACCAGGAGAGAAATTTGTTGATGTT 478

	AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
	TITLE	Functional annotation of a full-length mouse cDNA collection
	JOURNAL	Nature 409, 685-690 (2001)
	REFERENCE	
	AUTHORS	5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
	TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
	JOURNAL	Nature 420, 563-573 (2002)
	REFERENCE	6 (bases 1 to 1376)
	AUTHORS	Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saichou,H., Sakai,C., Sakai,X., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
	TITLE	Direct Submission
	JOURNAL	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]
	COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.jp/ URL:http://fantom.gsc.riken.jp/ Location/Qualifiers 1..1376 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM DB:E330039P08" /db_xref="taxon:10090" /clone="E330039P08" /sex="female" /tissue_type="ovary" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="2 days pregnant adult" 21..1286 /note="unnamed protein product; galactosidase, alpha (MGD GI:1347344, GB NM_013463, evidence: BLASTN, 99%, match=1373) putative" /codon_start=1 /protein_id="BAC35819.1" /db_xref="GI:26344325"
	CDS	/translation="MAKRLSRDRTLVCELALCPALVFWLSILGVRLDNLGLARTPTM GWLWRFMCNLDQCEPDACISQLFMQAEMLWSDGRDAGDYICIDCDNWAPER DSKGVLGLIYHNSKYLGLIGIVADGNKTCCAGFPGSFGSYDDAQT PDWGVDLLKDPGCHDSVLENGKYMALNRGTSIVYSWEFLYLPFFHKPNY TDIOYCNHRNFDDVDSWSIEINILSWTVYOKEIVEVAGPSNDPLMTVGCFR KENHQVTQMALWAIPAFLMSDLROIISOAKALLONKDVIAINODPLGKGVCYFR KENHLWEMERPLUSLNAAVAVRNLQEIGCGPCFYTIQISSLOGRLACNPGCITQLLEPR KVHLGFIEWTUTLKTRVNPSTGTVPFLRLR" 1358..1363 /note="putative" 1376 /note="putative"
	FEATURES	polyA_signal polyA_site ORIGIN

Query Match 71.5%; Score 897.2; DB 4; Length 1376;
Best Local Similarity 82.2%; Pred. No. 2.4e-257;
Matches 1031; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```
QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCGCTGCGCGCTTGGCGCTTCCCTCCCTGGCC 60
DB 27 ATGAAGCTTTTGACGAGATACCGCGCTGGTCTGTGAGCTTGGCGCTTGTCCCTGGCT 86
QY 61 CTGGTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATGGCAAGGACGCT 120
DB 87 TTAGTTTCTGGAGCAATCTTGGGGCTCAGAGCAATGGCAATGGCTTGGCGCGGACTCT 146
QY 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCTATGTGCAACCTTCACTGCCAGGAGACCA 180
DB 147 ACTATGGGCTGGCTGCAATGGGAACGTTTCTATGTGCAACCTTCACTGCCAGGAGACCT 206
QY 181 GATTCTGTCATCAGTGAAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
DB 207 GATGCTGTCATAGTGAAGCACTGTTTCATGCAAGTGCAGAGCTCATGGTCTCTGATGC 266
QY 241 TGAAGGATCGAGTTATGATGATCTCTGTCATGATGATGATGATGATGATGATGATGATG 300
DB 267 TGGCGGATCGAGTTATGATGATCTCTGTCATGATGATGATGATGATGATGATGATGATG 326
QY 301 GATTGAGAGCGAGACTTCAGGCGAGACCTTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
DB 327 GATTCAAGGGCGAGCTTCAGGCGAGATCCCAACGCTTTCCTAGTGGGATCAACACCTC 386
QY 361 GCTAATATCTTCACAGCAAGAGCTGAAGCTAGGAGTTATGCAAGATGTGGAATATAA 420
DB 387 GCAATATCTTCACAGCAAGAGATGAACTAGGAGTTATGCAAGATGTGGAATATAA 446
QY 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATCTATGACATGATGATGATGATGATGATGAT 480
DB 447 ACCTGTGAGGTTTCCCGGAGTTTGGATCTATGACATGATGATGATGATGATGATGATGAT 506
QY 481 GACTGGGAGTAGATCTGTAATATTTGATGATGATGATGATGATGATGATGATGATGATG 540
DB 507 GACTGGGCGTAGATCTGTAATATTTGATGATGATGATGATGATGATGATGATGATGATG 566
QY 541 GCAGATGTTTATAGCACATGTCCTGGCCCTGAATAGGAGCTGCGAGAGCATGTTGTAC 600
DB 567 GAGAAATGTTTATAGTACATGGGCTTGGCCCTTGAACAGGACAGCGCGAAGCATGTATAC 626
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660
DB 627 TCCTGTGAGTGGCCACTTTATTTGAGACCTTTTCATAAGCCCAATATATACAGATATCCA 686
QY 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATGATGATGATGATGATGATGATGATGAT 720
DB 687 TATTACTGCAATCAITGGAGAAATTTTGTATGATGATGATGATGATGATGATGATGATGAT 746
QY 721 AGTATCTGAGCTGGACATCTTTTAAACGAGAGAAATGTTGATGATGATGATGATGATGATG 780
DB 747 AATATCTGTTGAGAGTGGTTTACCAGAGAGAGATGTTGAAGTCGCTGGACAGGC 806
QY 781 GGTGGAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
DB 807 AGCTGGAATGACCCAGACATGTTAGTATGGCAACTTTGGCCCTCAGTTGGACAGCAG 866
QY 841 GTAATCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
DB 867 GTGACACAGATGCGCTCTGGGCTATCATGGCGCTCCCTACTCATGTCCACAGCATCTG 926
QY 901 CGACATCAGCCTCAAGCCAAAGCTCTCTTCAGATAGCACTTAATGCGCATCAAT 960
DB 927 CGCAAAATCAGCTCTCAAGCCAAAGCTCTCTTCAGATAGCACTTAATGCGCATCAAC 986
QY 961 CAGGACCCCTTGGCGAAGCAAGGATACCACTTTAGACAGGAGACCAACTTTGAAGTGTGG 1020
DB 987 CAGACCCCTTGGCGAAGCAAGGCTACTGTTTCAGAAAGGAAACCACTTGAAGTGTGG 1046
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCCGCGAGGAGATTGGT 1080
```

```
DB 1047 GAACGGCCACTCTCCCAACCTAGCCTGGCTGTGGCTGTGAGAACCTGCGAGGAGATTGGT 1106
QY 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTGAATCCT 1140
DB 1107 GGACCTTGCTTATACCATCCAGATTTCTTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1166
QY 1141 GCTGCTTCAATCAGACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTTATGAATGACT 1200
DB 1167 GCGTGCATCACTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGACT 1226
QY 1201 TCAAGGTTAAGAGTCACTAAATCCACAGGCACTGTTTGTCTTCAAGCTATAA 1254
DB 1227 TTGACCTTAAAACTCGAGTAAACCCCTCAGGCACCTGTTTGTCTTGGTAGAA 1280
```

RESULT 12

AK040081

LOCUS

DEFINITION

AK040081 2962 bp mRNA linear HTC 03-APR-2004
Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
library, clone:A430057F16 product:galactosidase, alpha, full insert
sequence.

ACCESSION

AK040081

VERSION

1 GI:263333580

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus

ORGANISM

Mus musculus

REFERENCE

1

Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

10345636

REFERENCE

2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

11042159

REFERENCE

3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

11076861

REFERENCE

4

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5

The PANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 2962)

REFERENCE

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,

Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,

Hori, P., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohse, N.,

VERSION	CO545623.1	GI:50567117
KEYWORDS	EST.	
SOURCE	Macaca mulatta (rhesus monkey)	
ORGANISM	Macaca mulatta	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Cercopithecidae; Cercopithecinae; Macaca.	
AUTHORS	1. (bases 1 to 922) Magnes, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agy, M.B., Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and Iadonato, S.P.	
TITLE	Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human	
JOURNAL	Genome Biol. 6 (7), R60 (2005)	
PUBMED	15998449	
COMMENT	Contact: C. Magnes Illumigen Biosciences Inc. 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA Tel: 2063780400 Fax: 2063780408 Email: cmagnes@illumigen.com Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof. Michael Katze Lab at University of Washington DNA Sequencing: Illumigen Biosciences Inc. For further information, see http://www.macaque.org PCR Primers FORWARD: CCCTCAATAAGGGACAAAA BACKWARD: CACTATAGGCGAATTGGGTA Insert Length: 922 Std Error: 0.00 Plate: C1000111 row: E Column: 08 Seq primer: CCCTCAATAAGGGACAAAA POLYA-No.	
FEATURES	Location/Qualifiers	
source	1. 922 /organism="Macaca mulatta" /mol_type="mRNA" /strain="Indian" /db_xref="taxon:9544" /clone="IBIUV:23923" /sex="female" /dev_stage="adult" /lab_host="Electromax DH10B" /clone_lib="Katze_WMLV" /note="Organ: liver; Vector: pDONR 222; Site 1: BarG I; Site 2: BarG I; Created from CloneMiner cdNA Library Construction kit (Catalog #18249-029)"	

Michael Katze Lab at University of Washington DNA Sequencing:
 Illumigen Biosciences Inc. For further information, see
<http://www.macaque.org>
 PCR Primers

FORWARD: CCTCTACTAAAGGACAAAA
 BACKWARD: CACTATAGGCGAATGGTGA
 Insert length: 1133 Std Error: 0.00
 Plate: CL000111 row: E column: 08
 Seq primer: CCTCTACTAAAGGACAAAA
 POLYA=No.

FEATURES

Location/Qualifiers

source

1..1133
 /organism="Macaca mulatta"
 /mol_type="mRNA"
 /strain="Indian"
 /db_xref="taxon:9544"
 /clone="IBIUM:23923"
 /sex="female"
 /dev_stage="adult"
 /lab_host="Electromax DH10B"
 /clone_lib="Katze MMLV"
 /note="Organ: liver; Vector: pDONR 222; Site 1: BerG I;
 Site 2: BerG I; Created from CloneMiner cDNA Library
 Construction kit (catalog #18249-029)"

ORIGIN

Query Match 70.3%; Score 881.8; DB 7; Length 1133;
 Best Local Similarity 94.6%; Pred. No. 9.3e-253;
 Matches 924; Conservative 0; Mismatches 52; Indels 1; Gaps 1;

QY 264 CCTCTGATGATGACGTGTTGGATGCTCCCAAGAGATTACAGAGGACAGCTTCAGGC 323
 DB 1041 CCTCTGCCCCGGGACTTTTGGAGATTCCCAAGAGACTCGAAGCGCCCTTCGGGC 982

QY 324 AGACCTCTCAGCGCTTCTCATGGGATTCCGACGTAGCTAAATTATGTTCCACAGCAAGG 383
 DB 981 GGGCGGTATGTTTCTTCATGGGATTGCCAGTTAGCTAAATTATGATTAACAGCAAGG 922

QY 384 ACTGAAGCTAGGATTTATCAGATGTTGAAATATAAACCTGCGAGGCTTCCTCGGAG 443
 DB 921 AATGAAATAGGATTTATGACAGCTTGGAAATATAAACCTGCGAGGCTTCCTCGGAG 862

QY 444 TTTTGGATACAGACATTGATGCCAGACTTGTGCTGACTGGGGAGTACATCTGCTAAA 503
 DB 861 TTTTGGATACATGACATTGATGCCAGCTTGTGCTGACTGGGGAGTACATCTGCTAAA 802

QY 504 ATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTGGCAGATGGTTTAAAGCACATGTC 563
 DB 801 ATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTGGCAGATGGTTTAAAGCACATGTC 742

QY 564 CTTGGCCCTGAATAGGACTGGCAGACATGTTGACTCTCTGTGAGTGGCCTCTTTATAT 623
 DB 741 GTTGGCCCTGAATAGGACTGGCAGACATGTTGACTCTCTGTGAGTGGCCTCTTTATAT 682

QY 624 GTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGACGACTGCTGCAATCACTGGCCAAA 683
 DB 681 GTGGCCCAATTTCAAAGCCCAATTATACAGAAATCCGACGACTGCTGCAATCACTGGCCAAA 622

QY 684 TTTTGTGATGATGATGATCTCTGAAAAGTATAAAGAGTATCTTTGACCTGGACATCTTT 743
 DB 621 TTTTGTGATGATGATGATCTCTGAAAAGTATAAAGAGTATCTTTGACCTGGACATCTTT 562

QY 744 TAACAGAGAGAAATTTGATGTTGTGACACAGGGGGTTGGAATGACCCAGATATGTT 803
 DB 561 TAACAGAGAGAAATTTGATGTTGTGACACAGGGGGTTGGAATGACCCAGATATGTT 502

QY 804 AGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCTCTGGC 863
 DB 501 AGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAAGTAACTCAGATGGCCTCTGGC 442

QY 864 TATCATGGCTGCTCTTTTATTCATGTTCTAATGACTCCGACATCAGCCCTCAAGCAA 923
 DB 441 TATCATGGCTGCTCTTTTATTCATGTTCTAATGACTCCGACATCAGCCCTCAAGCAA 382

QY 924 AGCTCTCCTTCAGGATAAGGACGTAATTGCCATCAATCAGGACCCCTTTGGGCAAGCAAGG 983
 DB 381 AGCTCTCCTTCAGGATAAGGATGTAATTGCCATCAATCAGGACCCCTTTGGGCAAGCAAGG 322

QY 984 GTACCAAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGC 1043
 DB 321 GTACCAAGCTTTAGACAGGAGAGACAACTTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGC 262

QY 1044 CTGGGCTGTAGCTATGATATAAACCCGGCAGAGAGATTGGTGGACCTCGCTCTTATACCATCGC 1103
 DB 261 CTGGGCTGTAGCTATGATATAAACCCGGCAGAGATCGGTGGACCTCGCTCTTATACCATCGC 202

QY 1104 AGTTTCTTCCCTGGGTAAAGGAGTGCCTGTAATCTGCTGCTTATCATCACAGAGCTCCT 1163
 DB 201 GGTGTCTTCCCTGGGTAAAGGAGTGCCTGTAATCTGCTGCTTATCATCACAGAGCTCCT 142

QY 1164 CCCTGTGAAAGGAAGCTAGGCTTCTATGAATCGACTTCAAGGTTAAGAGTCACTATAA 1223
 DB 141 CCCTG-GAAAGGAAGCTAGGCTTCTATGAATCGACTTCAAGGTTAAGAGTCACTATAA 83

QY 1224 TCCACAGGACATGTTT 1240
 DB 82 TCCACAGGACATGTTT 66

RESULT 15

AL575861/c

LOCUS

DEFINITION

AL575861 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA

Clone CS0D1067YJ01 3-PRIME, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1. (bases 1 to 1071)
 Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 On Feb 16, 2001 this sequence version replaced gi:31314157.

Contact: Genoscope - Centre National de Sequencage

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

10506.r

For more information about this cluster, see

<http://www.genoscope.cns.fr/cdna?s=CS0D1067CE01NFI&c=10506.r>.

FEATURES

source

1..1071
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1067YJ01"
 /tissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 69.8%; Score 874.8; DB 1; Length 1071;
 Best Local Similarity 91.6%; Pred. No. 1.1e-250;
 Matches 979; Conservative 20; Mismatches 61; Indels 9; Gaps 8;

Search completed: December 31, 2005, 21:00:26
Job time : 5621 secs

```
QY 191 TCAGTGAGAGCTCTTCATGAGATGCGAGAGCTCATGGTCTCAGAA-GGCTGGAAGGAT 249
Db 1071 TCAGTGAGAGCTC-TCATGNANATGNCAGANCTCATGNTCTCAGAACGCTGAAGNAT 1013

QY 250 GCAGGTTATGATGACCTCTGC--ATTGATGACTGTTGGATGGCT-CCCCAAGAGATTCA 306
Db 1012 GCAGTNTATGAGTACCTCTGCAATTCATGACTGTTGNATGNTCCCCCAAGAGATNCA 953

QY 307 GAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTAGCTAAT 366
Db 952 GAAGNCACACTNCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATNCGCCAGCTAGCTAAT 893

QY 367 TATGTTCAAGCAAGCACTGAAGCTAGGATTTATCCAGATGTTGGAATAAACCCTGC 426
Db 892 TATGCTCACAGCAAGCACTGAAGCTA-GGATTTATGCAGATGTT-GAATAAACCCTGC 835

QY 427 GCAGGCTTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCTGACTGG 486
Db 834 GCAGGCNTCCCTGGGAGTTTGGATCTACGACATTTGATGCCAGACCTTTGCTAACTGG 775

QY 487 GGAATAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTTGGCAGAT 546
Db 774 GGAATAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATCTVGCAGAT 715

QY 547 GGTATTAAGCACATGTCCTTTGGCCCTGAATAGGACTGGCAGAAAGCAATGTGTACTCCTGT 606
Db 714 GATWATAAGCACATGTCCTTTGGCCCTGAATAGAACTAGCAGAAAGCAATGTGTACACCTAT 655

QY 607 GAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCAAT-TATACAGAAATCCGACAGTA 665
Db 654 VAGTRACCTCACTATATCTGCCCTTMYCAAAGMCCAAATWTATACAAAAATCCGACAGTA 595

QY 666 CTGCAATCACTGGCGAAATTTGCTGACATTTGATGATTTCTCGAAAAAGTATAAAGAGTAT 725
Db 594 CTGCAATMACTGGCAAAATTTAGCTGACATTTGATGATTTCTGAAAAGTATAAAGAGTAT 535

QY 726 CTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACCAGGGGGTTG 785
Db 534 CTTGGACTGGACATCTWTTTAAACAGAGAAAGAAATTTGATGTTGCTGGACCAGGGGGTTG 475

QY 786 GAATGACCCAGATATGTTAGTATGTCGCACTTTGGCTCAGCTGGAATCAGCAAGTAAC 845
Db 474 GAATGACCCAGATATGTTAGTATGTCGCACTTTGGCTCAGCTGGAATCAGCAASTMAC 415

QY 846 TCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTCAATGACCTCCGACA 905
Db 414 TCAGAT-GCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTCAATGACCTCCGACA 356

QY 906 CATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAATCAGGA 965
Db 355 CATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAATCAGGA 296

QY 966 CCCCTTGGGCAAGCAGGGGTACAGCTTAGACAGGGAGACAACTTTGAAGTGTGGAAAG 1025
Db 295 CCCCTTGGGCAAGCAGGGGTACCRGCTTAGACAGGGAGACAACTTTGAAGTGTGGAAAG 236

QY 1026 ACCTCTCTCAGGCTTAGCCTGGGCTGATGATGATAAACCAGGAGAGATTGTTGGACC 1085
Db 235 ACCTCTCTCAGGCTTAGCCTGGGCTGATGATGATAAACCAGGAGAGATTGTTGGACC 176

QY 1086 TCGCTCTTATACCATCGAGTTGTTCCCTGGGTAAAGAGTGGCCTGTAAATCCTGCCTG 1145
Db 175 TCGCTCTTATACCATCGAGTTGTTCCCTGGGTAAAGAGTGGCCTGTAAATCCTGCCTG 116

QY 1146 CTTTCATCACAGCTCTCCTGTGAAAAGAAAGCTAGGGTTCTATGAATGGAATTCAG 1205
Db 115 CTTTCATCACAGCACTCCNKGTTGAAAAGAAAGCTAGGGTTCTATGAATGGAATTCAG 56

QY 1206 GTTAAGAGTACATTAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 55 GTTAAGAGTACATTAATCCCNACAGGCACTGTTTGGCTTCAGCTAGAA 7
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 17:12:51 ; Search time 261 Seconds
(without alignments)
8540.469 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgaggaccagca.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP_COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE_COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1254	100.0	1254	3	US-09-993-059-15
2	1254	100.0	1254	3	US-10-103-327-15
3	1252.4	99.9	1266	3	US-09-993-059-11
4	1252.4	99.9	1266	3	US-10-103-327-11
5	1252.4	99.9	1278	3	US-09-993-059-7
6	1252.4	99.9	1278	3	US-10-103-327-7
7	1252.4	99.9	1284	3	US-09-993-059-13
8	1252.4	99.9	1284	3	US-10-103-327-13
9	1252.4	99.9	1290	3	US-09-993-059-3
10	1252.4	99.9	1290	3	US-10-103-327-3
11	1252.4	99.9	1296	3	US-09-993-059-9
12	1252.4	99.9	1296	3	US-10-103-327-9
13	1252.4	99.9	1308	3	US-09-993-059-5
14	1252.4	99.9	1308	3	US-10-103-327-5
15	1252.4	99.9	1343	3	US-08-928-881-18
16	1252.4	99.9	1343	3	US-09-543-921-18
17	1252.4	99.9	1343	3	US-09-266-014-3
18	1252.4	99.9	1343	3	US-09-491-759-18
19	1252.4	99.9	1393	2	US-07-602-824A-1
20	1252.4	99.9	1393	2	US-07-983-451-1
21	1252.4	99.9	1393	2	US-08-261-577-6
22	1252	99.8	1272	3	US-09-993-059-17
23	1252	99.8	1272	3	US-10-103-327-17
24	1252	99.8	8234	3	US-09-626-127-14

25	1252	99.8	8234	3	US-09-993-059-34	Sequence 34, Appl
26	1252	99.8	8234	3	US-10-103-327-34	Sequence 34, Appl
27	1217.6	97.1	1304	9	5179023-3	Patent No. 5179023
28	1212.6	96.7	1233	3	US-09-993-059-21	Sequence 21, Appl
29	1212.6	96.7	1233	3	US-10-103-327-21	Sequence 21, Appl
30	1212	96.7	1215	3	US-09-993-059-19	Sequence 19, Appl
31	1212	96.7	1215	3	US-10-103-327-19	Sequence 19, Appl
32	1176.4	93.8	1214	3	US-09-023-655-1001	Sequence 1001, Ap
33	1164	92.8	1164	3	US-09-626-127-13	Sequence 13, Appl
34	1164	92.8	1164	3	US-09-993-059-33	Sequence 33, Appl
35	1164	92.8	1164	3	US-10-103-327-33	Sequence 33, Appl
36	1159.4	92.5	1161	3	US-09-176-666-52	Sequence 52, Appl
37	1159.4	92.5	1164	3	US-09-176-666-51	Sequence 51, Appl
38	1159.4	92.5	1167	3	US-09-176-666-50	Sequence 50, Appl
39	1159.4	92.5	1170	3	US-09-176-666-49	Sequence 49, Appl
40	1159.4	92.5	1173	3	US-09-176-666-48	Sequence 48, Appl
41	1159.4	92.5	1176	3	US-09-176-666-47	Sequence 47, Appl
42	1159.4	92.5	1179	3	US-09-176-666-46	Sequence 46, Appl
43	1159.4	92.5	1182	3	US-09-176-666-45	Sequence 45, Appl
44	1159.4	92.5	1188	3	US-09-176-666-44	Sequence 44, Appl
45	1159.4	92.5	1194	3	US-09-176-666-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 08010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Query Match	100.0%;	Score 1254;	DB 3;	Length 1254;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1254;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60	
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTGGCC	60	
QY	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATTGGCAAGGACGCCT	120	
Db	61	CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGATTGGCAAGGACGCCT	120	
QY	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA	180	
Db	121	ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA	180	
QY	181	GATTCTGTCATGATGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC	240	
Db	181	GATTCTGTCATGATGAGAAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC	240	
QY	241	TGGAAGGATCGAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300	
Db	241	TGGAAGGATCGAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA	300	

```
301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
301 GATTGAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
361 GCTAATTATGTTACAGCAAGAGGAGCTGAGCTAGGAGATTTATGACAGATGTTGGAATAAA 420
361 GCTAATTATGTTACAGCAAGAGGAGCTGAGCTAGGAGATTTATGACAGATGTTGGAATAAA 420
421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGATGATGATGCCAGACCTTTTGCT 480
421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGATGATGATGCCAGACCTTTTGCT 480
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAAA 540
481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAAA 540
541 GCAGATGGTTATAGACACATGCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
541 GCAGATGGTTATAGACACATGCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
721 AGTATCTGACCTGGAGATCTTTTAAACAGAGAGAAATTTGTCATGTTGCTGACACAGG 780
721 AGTATCTGACCTGGAGATCTTTTAAACAGAGAGAAATTTGTCATGTTGCTGACACAGG 780
781 GGTGGAATGACCCAGATATGTTAGTATGTTGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
781 GGTGGAATGACCCAGATATGTTAGTATGTTGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
841 GTAACTGAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATGATGCTAATGACCTC 900
841 GTAACTGAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATGATGCTAATGACCTC 900
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
1021 GAACGACCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATGATATGATATGATATGAT 1080
1021 GAACGACCTCTCAGGCTTAGGCTGGGCTGTAGCTATGATATGATATGATATGATATGAT 1080
1081 GGACCTGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTCTGATTCCT 1140
1081 GGACCTGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGAGAGTGGCTCTGATTCCT 1140
1141 GCCTGCTTTCATCAGACGCTCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGAGCT 1200
1141 GCCTGCTTTCATCAGACGCTCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGAGCT 1200
1201 TCAAGGTTAAGAGTGCATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
1201 TCAAGGTTAAGAGTGCATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
```

RESULT 2

```
US-10-103-327-15
; Sequence 15, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
```

```
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Query Match          100.0%; Score 1254; DB 3; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCGCTTCCTGGCC 60
DB      1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCGCTTCCTGGCC 60
QY     61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGGATGGCAAGACGCT 120
DB     61 CTCGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGGATGGCAAGACGCT 120
QY    121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTCGCTGCCAGGAGAGCC 180
DB    121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATCTGCAACCTTCGCTGCCAGGAGAGCC 180
QY    181 GATTCTCTGCATCAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
DB    181 GATTCTCTGCATCAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY    241 TGAAGAGATGAGGTTATGAGTACCTCTGCATGATGAGCTGTTGGATGGCTCCCAAGA 300
DB    241 TGAAGAGATGAGGTTATGAGTACCTCTGCATGATGAGCTGTTGGATGGCTCCCAAGA 300
QY    301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCT 360
DB    301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCT 360
QY    361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGAGATTTATGACAGATGTTGGAATAAA 420
DB    361 GCTAATTATGTTTCACAGCAAGAGCTGAAGCTAGGAGATTTATGACAGATGTTGGAATAAA 420
QY    421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGATGATGATGCCAGACCTTTTGCT 480
DB    421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACGATGATGATGCCAGACCTTTTGCT 480
QY    481 GACTGGGAGCTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAAA 540
DB    481 GACTGGGAGCTAGATCTGCTAAATTTGATGTTGTTACTGACAGATTTGGAATAAA 540
QY    541 GCAGATGGTTATAGGACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
DB    541 GCAGATGGTTATAGGACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY    601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
DB    601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY    661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
DB    661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTTCTGGAAAGTATAAAG 720
QY    721 AGTATCTGACCTGGAGATCTTTTAAACAGAGAGAAATTTGTCATGTTGCTGACACAGG 780
```

Db 721 AGTATCTTGAGCTGGACATCTTTAAACAGAGAGAAATGTTGATGTTGCTGACCAAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGCTAAATGGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGCTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGATACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGATACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGATGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTGCATATAATCCACAGGACCTGTTTGTCTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTGCATATAATCCACAGGACCTGTTTGTCTCAGCTATATA 1254

RESULT 3

US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GABER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11

Query Match 99.9%; Score 1252.4; DB 3; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTGGCC 60
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTTCTGGCC 60
Qy 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCTAGGCAATGGATGGCAAGAGCGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCTAGGCAATGGATGGCAAGAGCGCT 120
Qy 121 ACCATGGCTGCTGACCTGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180
Db 121 ACCATGGCTGCTGACCTGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180

Db 121 ACCATGGCTGCTGACCTGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180
Qy 181 GATTCTTCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTTCATCAGTGGAGAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGAAGGATGACAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGACAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGAGCAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGCGATTTCGCGAGCTA 360
Db 301 GATTTCAGAGGAGCAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGCGATTTCGCGAGCTA 360
Qy 361 GCTAAATATGTTTACAGCAAGAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAAA 420
Db 361 GCTAAATATGTTTACAGCAAGAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACAGATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACAGATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTACATCTGCTAAATTTGATGTTGTTTACTGTGACACTTTGGAAAAATTG 540
Db 481 GACTGGGAGTACATCTGCTAAATTTGATGTTGTTTACTGTGACACTTTGGAAAAATTG 540
Qy 541 GCAGATGGTTTAAAGCACATGCTCCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 600
Db 541 GCAGATGGTTTAAAGCACATGCTCCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGGGAATTTGCTGACATTTGATGATTCCTCGAAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGGGAATTTGCTGACATTTGATGATTCCTCGAAAAAGTATAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGTCGACCAAGG 780
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGTCGACCAAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGCTAAATGGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGACGCTAAATGGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGATACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGATACCAAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGACAGGATGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTGCATATAATCCACAGGACCTGTTTGTCTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTGCATATAATCCACAGGACCTGTTTGTCTCAGCTATATA 1254

```
RESULT 4
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11

Query Match          99.9%; Score 1252.4; DB 3; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
DB 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAGGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAAGA 300
DB 241 TGGAAAGGATGCAGGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAAGA 300
QY 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAATAAAA 420
DB 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAATAAAA 420
QY 421 ACCTGGCGAGGCTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCGAGGCTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTCGAAAAATTTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTCGAAAAATTTG 540
QY 541 GCAGATGGTTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACCATTTGTATC 600
DB 541 GCAGATGGTTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGGAATTTTGTGATGATGATGATTTCTCGGAAAGATATAAG 720
DB 661 CAGTACTGCAATCACTGGCGGAATTTTGTGATGATGATGATTTCTCGGAAAGATATAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAGAAATTTGTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTTGGGCTCAGCTGGAAATCAGCAA 840
DB 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTTGGGCTCAGCTGGAAATCAGCAA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATCATGCTAATGACCTC 900
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATCATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
DB 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATTTGGT 1080
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCGCAGGAGATTTGGT 1080
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT 1140
QY 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGGACT 1200
DB 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAGCTAGGGTTCTATGATGGACT 1200
QY 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTATAA 1254
DB 1201 TCAAGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTATAA 1254

RESULT 5
US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Query Match          99.9%; Score 1252.4; DB 3; Length 1278;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
QY 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
DB 61 CTGGTTCTCGGACATCCCTGGGCTAGAGCACTGCAACATGATTTGGCAAGAGCGCT 120
QY 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
DB 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
DB 181 GATTCCTGCATAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 241 TGGAAAGGATGCAGGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAAGA 300
DB 241 TGGAAAGGATGCAGGTTATGATGATACCTCTGCATTGATGACTGTTGGATGGCTCCCAAAGA 300
QY 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
DB 301 GATTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAATAAAA 420
DB 361 GCTAAATATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGAGATGTTGGAATAAAA 420
QY 421 ACCTGGCGAGGCTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
DB 421 ACCTGGCGAGGCTCCCTGGGAGTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTCGAAAAATTTG 540
DB 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTTGTACTGTGACAGTTCGAAAAATTTG 540
QY 541 GCAGATGGTTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACCATTTGTATC 600
DB 541 GCAGATGGTTTATAGCAATGCTTCCTGGCCCTGAAATAGGACTGGCAGAACCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
DB 601 TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
```

Db 1 ATGCAGCTGAGAACCCAGAACTATACATCTGGGCTGCGCGCTTGGCTTCCCTGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGAGCGCTTATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGAGCGCTTATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGAAGGATGCAAGTATAGTACCTCTGCAATGATGACTGTGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCAAGTATAGTACCTCTGCAATGATGACTGTGATGGCTCCCAAGA 300
Qy 301 GATTGAGAGGAGACTTTCAGGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGACTTTCAGGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGCAATGATGCGCAATATAA 420
Db 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGCAATGATGCGCAATATAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGACATGATGCCAGACTTTGCT 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGACATGATGCCAGACTTTGCT 480
Qy 481 GACTGGGAGTAGACTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGAGTAGACTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGTTTAAAGACATGCTTGGCCCTGAATAGGACTGCGAGCAATGATGCTAT 600
Db 541 GCAGATGTTTAAAGACATGCTTGGCCCTGAATAGGACTGCGAGCAATGATGCTAT 600
Qy 601 TCCTGTGAGTGGCTCTTTATATATGCGCCCTTTCAAAAGCCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTTCTGGAAGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTGCTGCAATGATGATTTCTGGAAGATATAAG 720
Qy 721 AGTATCTTGACATGGAATCTTTTAAACGAGAGAAATGTTGATGTTGCTGACAGGG 780
Db 721 AGTATCTTGACATGGAATCTTTTAAACGAGAGAAATGTTGATGTTGCTGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCAACTTTGGCTCAGCTGGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGAATGGCAACTTTGGCTCAGCTGGAATCAGCA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATGATGCTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAGCCAAAGCTCTCTCAGGATGAGGAGCTTAATGCGCATCAAT 960
Db 901 CGACATCAGCCCTCAGCCAAAGCTCTCTCAGGATGAGGAGCTTAATGCGCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGCAGGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGCAGGAGCAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGGAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGGAGGAGATGGT 1080
Qy 1081 GGAACCTCGCTTATACCATGCGAGTTGCTTCCCTGGGTAAGAGTGGCTGTATCT 1140

Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTATCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy 1201 TCAAGCTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTATAA 1254
Db 1201 TCAAGCTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTAGAA 1254

RESULT 6

US-10-103-327-7
; Sequence 7, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7

Query Match 99.9%; Score 1252.4; DB 3; Length 1278;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGAACCCAGAACTATACATCTGGGCTGCGCGCTTGGCTTCCCTGGCC 60
Db 1 ATGCAGCTGAGAACCCAGAACTATACATCTGGGCTGCGCGCTTGGCTTCCCTGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGAACAATGATGGCAAGAGCGCT 120
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCAATGTGCAACCTTGACCTGCCAGGAGGCCA 180
Qy 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTGTCATCAGTGAGAGCTCTTCATGAGATGCGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGAAGGATGCAAGTATAGTACCTCTGCAATGATGACTGTGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCAAGTATAGTACCTCTGCAATGATGACTGTGATGGCTCCCAAGA 300
Qy 301 GATTGAGAGGAGACTTTCAGGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTGAGAGGAGACTTTCAGGAGACCTTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGCAATGATGCGCAATATAA 420
Db 361 GCTAATTATGTTTACAGCAAGGACTGAAGCTAGGATTTATGCAATGATGCGCAATATAA 420
Qy 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGACATGATGCCAGACTTTGCT 480
Db 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGACATGATGCCAGACTTTGCT 480


```
QY 481 GACTGGGAGTAGACTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db |||||||
QY 481 GACTGGGAGTAGACTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db |||||||
QY 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAAAGCAATTTGTATC 600
Db |||||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||||
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db |||||||
QY 661 CAGTACTGCAATCACATGGCGAAATTTTCTGCAATTTGATGATTTCTTGGAAAGTATAAAG 720
Db |||||||
QY 661 CAGTACTGCAATCACATGGCGAAATTTTCTGCAATTTGATGATTTCTTGGAAAGTATAAAG 720
Db |||||||
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db |||||||
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAATTTGGCCCTCAGCTGGATCAGCA 840
Db |||||||
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAATTTGGCCCTCAGCTGGATCAGCA 840
Db |||||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCAATGCTAATGACCTC 900
Db |||||||
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCAATGCTAATGACCTC 900
Db |||||||
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTAGGATAAGGACGTAATTTGCCATCAAT 960
Db |||||||
QY 961 CAGGACCCCTTGGGCAAGCAGGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db |||||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGCGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Db |||||||
QY 1021 GAACGACCTCTCTCAGGCTTAGCTGCGCTGTAGCTATGATAAACCGGACAGGAGATTGGT 1080
Db |||||||
QY 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
Db |||||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGCAAAAGGAGTGGCTTCTATGAATGGACT 1200
Db |||||||
QY 1141 GCCTGCTTATACACAGCTCTCCCTGTGCAAAAGGAGTGGCTTCTATGAATGGACT 1200
Db |||||||
QY 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db |||||||
QY 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTCAGCTATATA 1254
Db |||||||
```

RESULT 7

```
US-09-993-059-13
; Sequence 13, Application US/09993059
; Patent No. 688796
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-09-993-059-13
Query Match 99.9%; Score 1252.4; DB 3; Length 1284;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGCAAGCTCAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCGCTTGCCTTCCCTGGCC 60
Db 1 ATGCAAGCTCAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCGCTTGCCTTCCCTGGCC 60
QY 61 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
QY 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACTTTGACTGCCAGGAGAGCCA 180
QY 181 GATTCTGTCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
Db 181 GATTCTGTCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240
QY 241 TGAAGGATGCAGGTTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
Db 241 TGAAGGATGCAGGTTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
QY 301 GATTCAAGAGGACAGCTTTCAGGAGACCTTCAGGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGAGGACAGCTTTCAGGAGACCTTCAGGCGCTTCCTCATGGGATTCGCCAGCTA 360
QY 361 GCTAATTTATGTTTACACAGCAAGGACCTGAAGCTAGGAGATTTATGCAGATGTTGAAATATA 420
Db 361 GCTAATTTATGTTTACACAGCAAGGACCTGAAGCTAGGAGATTTATGCAGATGTTGAAATATA 420
QY 421 ACCTGCGCAGGCTTCCCTCGGAGTTTTGGATCTACGACATTCATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTCGGAGTTTTGGATCTACGACATTCATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGACTGCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Db 481 GACTGGGAGTAGACTGCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAAGCAATTTGTAC 600
Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAAGCAATTTGTAC 600
QY 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACATGGGAGAAATTTTGTGATGATTTCTTGGAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACATGGGAGAAATTTTGTGATGATTTCTTGGAAAGTATAAAG 720
QY 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGGACAGGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAATTTGGCCCTCAGCTGGAAATCAGCA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGTCGCAATTTGGCCCTCAGCTGGAAATCAGCA 840
QY 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCAATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCAATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGGAGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
```

```
Db 961 CAGGACCCCTTGGCAAGCAAGGTAACAGCTTAGACAGGAGACAACCTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTATGAATGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTATGAATGACT 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAAATCCCAAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAAATCCCAAGGCACTGTTTGGCTTCAGCTAGAA 1254

RESULT 8
US-10-103-327-13
; Sequence 13, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13

Query Match 99.98; Score 1252.4; DB 3; Length 1284;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Qy 61 CTGCTTTCTGGGACATCCCTGGGCTTAGAGCACTGACACATGGAATGGCAAGACGCT 120
Db 61 CTGCTTTCTGGGACATCCCTGGGCTTAGAGCACTGACACATGGAATGGCAAGACGCT 120
Qy 121 ACCATGGGCTGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
Db 121 ACCATGGGCTGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
Qy 181 GATTCTCTGATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGAAGAGATGCAAGTTATGATGATACCTCTGCAATGATGATGATGATGATGATGATGATGAT 300
Db 241 TGAAGAGATGCAAGTTATGATGATACCTCTGCAATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GATTTCAGAGGACAGCTTCAGGAGACCCCTCAGGCTTCTCATGGATTCGCCAGCTA 360
Db 301 GATTTCAGAGGACAGCTTCAGGAGACCCCTCAGGCTTCTCATGGATTCGCCAGCTA 360
```

```
Qy 361 GCTAAATTATGTTTCAAGCAAGAGCTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 420
Db 361 GCTAAATTATGTTTCAAGCAAGAGCTGAAGCTAGGGATTTATGAGATGTTGGAAATAAA 420
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACAGATGATGATGATGATGATGATGATGAT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACACAGATGATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAAATTG 540
Qy 541 GCAGATGGTTATTAAGCAATGCTTGGCCCTGAAATAGGACTGCGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGGTTATTAAGCAATGCTTGGCCCTGAAATAGGACTGCGCAGAGCAATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTTCTCGAAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCTGACATTTGATGATTTCTCGAAAAAGTATAAG 720
Qy 721 AGTATCTTGAATGAGCACTTTTAAACAGGAGAGAAATTTGATGATTTGTGTCGACAGGG 780
Db 721 AGTATCTTGAATGAGCACTTTTAAACAGGAGAGAAATTTGATGATTTGTGTCGACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 781 GGTGGAAATGACCCAGATATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Qy 841 GTAACTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTTATGATGATGATGATGAT 900
Db 841 GTAACTCAGATGCGCTCTGGGCTATCATGGCTGCTCTTTTATTTATGATGATGATGATGAT 900
Qy 901 CGACATCATGAGCTTCAAGCCAAAGCTCTCTTCAAGATTAAGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATGAGCTTCAAGCCAAAGCTCTCTTCAAGATTAAGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGAGCAAGGTAACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGAGCAAGGTAACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGGCTGTAGCTATGATAAACCAGGAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTATGAATGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTCTATGAATGACT 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAAATCCCAAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAAATCCCAAGGCACTGTTTGGCTTCAGCTAGAA 1254
```

```
RESULT 9
US-09-993-059-3
; Sequence 3, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
```

```
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-993-059-3

Query Match          99.9%; Score 1252.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGCAAGGACGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAAGCCCA 180
Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAAGCCCA 180
QY 181 GATTCTCGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTCGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 241 TGAAGGATGCAGGTTATGATGACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCAGGTTATGATGACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 301 GATTCAAGGAGAGACTTCAGGAGAGCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
Db 301 GATTCAAGGAGAGACTTCAGGAGAGCCCTCAGGCTTTCCTCATGGATTCGCCAGCTA 360
QY 361 GCTAATTTATCTCACAGCAAGAGACTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
Db 361 GCTAATTTATCTCACAGCAAGAGACTGAAGCTAGGAGATTTATGAGATGTTGGAATAAA 420
QY 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGGATCTAGAGATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGGATCTAGAGATGATGATGCCAGACCTTTGCT 480
QY 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG 540
QY 541 GCAGATGTTATAGACATGTCCTGGCCCTGAAATAGGACTGGCAGAAGCATTTGTATC 600
Db 541 GCAGATGTTATAGACATGTCCTGGCCCTGAAATAGGACTGGCAGAAGCATTTGTATC 600
QY 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATGATGATTCCTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTCTGACATGATGATTCCTGGAAAAGTATAAG 720
QY 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGAGAGAAATTTGATGTTGCTGGACACAGG 780
QY 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
```

```
QY 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTAATGCTAATGACCTC 900
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTAATGCTAATGACCTC 900
QY 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCCATCAAT 960
QY 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
QY 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT 1080
QY 1081 GGAACCTCGCTTATACCATCGCAGTTGCTTCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
Db 1081 GGAACCTCGCTTATACCATCGCAGTTGCTTCCTGGGTAAGGAGTGGCCTGTAATCCT 1140
QY 1141 GCCTGCTTCATCACAGACTCCCTCTGTGAAGGAAGCTAGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGACTCCCTCTGTGAAGGAAGCTAGGTTCTATGAATGGACT 1200
QY 1201 TCAAGCTTAAGAGTCAATAAATCCACAGGCACTGTTTTCCTCAGCTATAA 1254
Db 1201 TCAAGCTTAAGAGTCAATAAATCCACAGGCACTGTTTTCCTCAGCTATAA 1254

RESULT 10
US-10-103-327-3
; Sequence 3, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCES: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-10-103-327-3

Query Match          99.9%; Score 1252.4; DB 3; Length 1290;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
QY 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGCAAGGACGCT 120
Db 61 CTCGTTTCTTGGGACATCCCTGGGCTTAGAGCACTGGACAATGGCAAGGACGCT 120
QY 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAAGCCCA 180
Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTCACTGCCAGGAAGCCCA 180
QY 181 GATTCTCGCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
```

```
Db 181 GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGCGAGAGCTCATGGTCTCAGAGGC 240
Qy 241 TGAAGAGATGAGGTATAGTACTCTGCAATGATGACTGTGGATGGCTCCCAAAGA 300
Db 241 TGAAGAGATGAGGTATAGTACTCTGCAATGATGACTGTGGATGGCTCCCAAAGA 300
Qy 301 GATTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGATTCGCAGCTA 360
Db 301 GATTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGATTCGCAGCTA 360
Qy 361 GCTAATATATGTTACAGCAAGAGCTGAAGCTAGGGATTTATGAGATGTTGGAATAA 420
Db 361 GCTAATATATGTTACAGCAAGAGCTGAAGCTAGGGATTTATGAGATGTTGGAATAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGAGATTTGGATPACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTATAGCACATGCTTCCGCTCTGAATAGGACTGCGAGAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTATAGCACATGCTTCCGCTCTGAATAGGACTGCGAGAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATTAACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATTAACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGACATGGAATCTTTTAAACAGAGAGAAATTTGATGTTGTCGACACAGGG 780
Db 721 AGTATCTTGACATGGAATCTTTTAAACAGAGAGAAATTTGATGTTGTCGACACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGCAACTTTGGCTTCAGCTGGAATACGAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGCAACTTTGGCTTCAGCTGGAATACGAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGATACAGCTTTAGACAGGAGACAACTTTGAAAGTGTG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATATAACCGGCGAGGAGATGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATATAACCGGCGAGGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGAATCCCT 1140
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGAATCCCT 1140
Qy 1141 GCCTGCTTATCAACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTATCAACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTACATAAATCCCAAGGCACTGTTTTGCTTCAGCTATATA 1254
Db 1201 TCAAGGTTAAGAGTACATAAATCCCAAGGCACTGTTTTGCTTCAGCTATATA 1254
```

RESULT 11

US-09-993-059-9

```
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
```

```
Query Match 99.9%; Score 1252.4; DB 3; Length 1296;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCGCTTCCTGGCC 60
Qy 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTTGGCAAGACGCT 120
Db 61 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGAACAATGGATTTGGCAAGACGCT 120
Qy 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTCCAGGAAGAGCCA 180
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTCCAGGAAGAGCCA 180
Qy 181 GATTCTGTCATAGTGAGAAAGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Db 181 GATTCTGTCATAGTGAGAAAGCTTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 241 TGAAGAGATGACAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAAGA 300
Db 241 TGAAGAGATGACAGTTATGAGTACCTCTGCAATGATGACTGTGGATGGCTCCCAAAGA 300
Qy 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTTGCCAGCTA 360
Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTTGCCAGCTA 360
Qy 361 GCTAATATATGTTACAGCAAGAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAA 420
Db 361 GCTAATATATGTTACAGCAAGAGGACTGAAGCTAGGGATTTATGAGATGTTGGAATAA 420
Qy 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACTACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Db 481 GACTGGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTATAGCACATGCTTCCGCTCTGAATAGGACTGCGAGAAGCATTTGTGTAC 600
Db 541 GCAGATGGTTATAGCACATGCTTCCGCTCTGAATAGGACTGCGAGAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTAACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGCCCAATTAACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCACTGGCGCAATTTTGTGACATTTGATGATTTCTTGGAAAAGTATAAG 720
```

QY	721	AGTATCTTGGACTGGACATCTTTTAAACAGAGGAGAAATGTTGATGTTGCTCGACAGGG	780	Db	61	CTCGTTCTCGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGAGCGCT	120
Db	721	AGTATCTTGGACTGGACATCTTTTAAACAGAGGAGAAATGTTGATGTTGCTCGACAGGG	780	QY	121	ACATGGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA	180
QY	781	GTTTGGAAATGACCCAGATATGTAGTATGTTGCACTTTGGGCTCAGCTGGAATCAGCAA	840	Db	121	ACCATGGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA	180
Db	781	GTTTGGAAATGACCCAGATATGTAGTATGTTGCACTTTGGGCTCAGCTGGAATCAGCAA	840	QY	181	GATTCCTGCATCAGTGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGTCTCAGAAGGC	240
QY	841	GTAACCTCAGATGGGCTCTCGGGCTATCATGGCTGTCTCTTTATTTATGTTCTAATGACCTC	900	Db	181	GATTCCTGCATCAGTGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGTGTCTCAGAAGGC	240
Db	841	GTAACCTCAGATGGGCTCTCGGGCTATCATGGCTGTCTCTTTATTTATGTTCTAATGACCTC	900	QY	241	TGGAAGGATGCGAGTTTATGAGTACCTCTGCATTTGATGACTGTGTGGATGGCTCCCCAAAGA	300
QY	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT	960	Db	241	TGGAAGGATGCGAGTTTATGAGTACCTCTGCATTTGATGACTGTGTGGATGGCTCCCCAAAGA	300
Db	901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT	960	QY	301	GATTCAGAAAGCGAGCTTCAGGAGACCTCAGCGCTTTCCTCATGGATTCGCCAGCTA	360
QY	961	CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020	Db	301	GATTCAGAAAGCGAGCTTCAGGAGACCTCAGCGCTTTCCTCATGGATTCGCCAGCTA	360
Db	961	CAGGACCCCTTGGGCAAGCAAGGTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020	QY	361	GCTAATTATGTTTACAGCAAGGACTGGAAGCTAGGATTTATGAGATGTTGGAATATAA	420
QY	1021	GAAAGACCTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT	1080	Db	361	GCTAATTATGTTTACAGCAAGGACTGGAAGCTAGGATTTATGAGATGTTGGAATATAA	420
Db	1021	GAAAGACCTCTCTCAGGCTTAGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT	1080	QY	421	ACCTGCGCAGGCTTCCCTGGAGTTTGGATCTACGACATTCAGTATGATGCTTGGCT	480
QY	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGTAAAGGAGTGGCTGTATCTCT	1140	Db	421	ACCTGCGCAGGCTTCCCTGGAGTTTGGATCTACGACATTCAGTATGATGCTTGGCT	480
Db	1081	GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGTAAAGGAGTGGCTGTATCTCT	1140	QY	481	GACTGGGGAGTAGATCTGCTTAAATTTGATGTTGTTACGTGACAGTGTGGAAATTTG	540
QY	1141	GCTGCTTATACCATCGAGTTGCTTCCCTGGTAAAGGAGTGGCTGTATCTCT	1200	Db	481	GACTGGGGAGTAGATCTGCTTAAATTTGATGTTGTTACGTGACAGTGTGGAAATTTG	540
Db	1141	GCTGCTTATACCATCGAGTTGCTTCCCTGGTAAAGGAGTGGCTGTATCTCT	1200	QY	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAACATTTGTGATC	600
QY	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATATA	1254	Db	541	GCAGATGTTTATAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAAACATTTGTGATC	600
Db	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTATATA	1254	QY	601	TCTGTGAGTGGCTCTTTATATGTTGCTTCAAAAAGCCCAATTATACAGAAATCCGA	660
RESULT 12							
US-10-103-327-9							
; Sequence 9, Application US/10103327							
; Patent No. 6890748							
; GENERAL INFORMATION:							
; APPLICANT: GARGER, Stephen A.							
; APPLICANT: TURPEN, Thomas H.							
; APPLICANT: KUMAGAI, Monto H.							
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN							
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION							
; FILE REFERENCE: 008010087CPUS06							
; CURRENT APPLICATION NUMBER: US/10/103,327							
; PRIOR FILING DATE: 2002-03-20							
; PRIOR APPLICATION NUMBER: US/09/993,059							
; PRIOR FILING DATE: 2001-11-13							
; NUMBER OF SEQ ID NOS: 37							
; SOFTWARE: FastSeq for Windows Version 4.0							
; SEQ ID NO 9							
; LENGTH: 1296							
; TYPE: DNA							
; ORGANISM: Homo sapiens							
; FEATURE:							
; NAME/KEY: CDS							
; LOCATION: (1)... (1296)							
US-10-103-327-9							
Query Match							
Best Local Similarity 99.9%; Score 1252.4; DB 3; Length 1296;							
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;							
QY	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGTGCTCGGCC	60				
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCGTGCTCGGCC	60				
QY	61	CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGGATTGGCAAGGACGCT	120				

Job time : 266 secs

```
|||||
Db 216 GATTCCTGCAATCAGTGAGAGCTCTTCATGGAGATGCGAGGCTCATGGTCTCAGAAGGC 275
Qy
Db 241 TGGAAAGATGACAGCTTATGAGTACTCTGCAATTCATGACTGTTGGATGGCTCCCCAAAGA 300
Qy
Db 276 TGGAGGATGACAGGTTATGATGACTCTGCAATTCATGATGACTGTTGGATGGCTCCCCAAAGA 335
Qy
Db 301 GATTCAGAGGCGAGACTTCAGGCGAGACCCCTCAGCGCTTTCTCATNGGGATTGCGCCAGCTA 360
Qy
Db 336 GATTCAGAGGCGAGACTTCAGGCGAGACCCCTCAGCGCTTTCTCATNGGGATTGCGCCAGCTA 395
Qy
Db 361 GCTAAATTTATCTTCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy
Db 396 GCTAAATTTATCTTCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 455
Qy
Db 421 ACCTGCCAGGCTTCCCTGGGAGTTTGGGATATCTACGACATTCATGCCAGACCTTTGCT 480
Qy
Db 456 ACCTGCCAGGCTTCCCTGGGAGTTTGGGATATCTACGACATTCATGCCAGACCTTTGCT 515
Qy
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTG 540
Qy
Db 516 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTTG 575
Qy
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGNAATAGGACTGGCAGAGCATTTGTAC 600
Qy
Db 576 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTAC 635
Qy
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy
Db 636 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 695
Qy
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAAATATAAAG 720
Qy
Db 696 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAAATATAAAG 755
Qy
Db 721 AGTATCTTGCACTGGACATCTTTTACCAGGAGAGATTTGTGATGTTGCTGGACCAGGG 780
Qy
Db 756 AGTATCTTGCACTGGACATCTTTTACCAGGAGAGATTTGTGATGTTGCTGGACCAGGG 815
Qy
Db 781 GGTGGAATGACCCAGATATGTTAGTATGATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy
Db 816 GGTGGAATGACCCAGATATGTTAGTATGATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 875
Qy
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy
Db 876 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 935
Qy
Db 901 CGACACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCGCATCAAT 960
Qy
Db 936 CGACACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCGCATCAAT 995
Qy
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGCAGGAGAGCAACTTTGAAAGTGTGG 1020
Qy
Db 996 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGCAGGAGAGCAACTTTGAAAGTGTGG 1055
Qy
Db 1021 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy
Db 1056 GAACGACCTCTCTCAGGCTTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1115
Qy
Db 1081 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTATCCT 1140
Qy
Db 1116 GGACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTATCCT 1175
Qy
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy
Db 1176 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1235
Qy
Db 1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGACCTGTTTTGCTTCAGCTATAA 1254
Qy
Db 1236 TCAAGGTTAAGAGTCAATAAATCCCAAGGACCTGTTTTGCTTCAGCTAGAA 1289
```

Search completed: December 31, 2005, 21:04:52

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 19:26:56 ; Search time 1160 Seconds
(without alignments)
8939.483 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgaggaccaga.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA_Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1254	100.0	1254	3	US-09-993-059-15
2	1254	100.0	1254	5	US-10-103-327-15
3	1254	100.0	1254	7	US-10-602-219-15
4	1254	100.0	1254	7	US-10-602-220-15
5	1254	100.0	1254	8	US-10-851-388-15
6	1254	100.0	1254	9	US-10-984-389-15
7	1252.4	99.9	1266	5	US-09-993-059-11
8	1252.4	99.9	1266	5	US-10-103-327-11
9	1252.4	99.9	1266	7	US-10-602-219-11
10	1252.4	99.9	1266	7	US-10-602-220-11
11	1252.4	99.9	1266	8	US-10-851-388-11
12	1252.4	99.9	1266	9	US-10-984-389-11
13	1252.4	99.9	1278	3	US-09-993-059-7
14	1252.4	99.9	1278	5	US-10-103-327-7
15	1252.4	99.9	1278	7	US-10-602-219-7
16	1252.4	99.9	1278	7	US-10-602-220-7
17	1252.4	99.9	1278	8	US-10-851-388-7
18	1252.4	99.9	1278	9	US-10-984-389-7
19	1252.4	99.9	1284	3	US-09-993-059-13
20	1252.4	99.9	1284	5	US-10-103-327-13
21	1252.4	99.9	1284	7	US-10-602-219-13
22	1252.4	99.9	1284	7	US-10-602-220-13
23	1252.4	99.9	1284	8	US-10-851-388-13

24	1252.4	99.9	1284	9	US-10-984-389-13	Sequence 13, Appl
25	1252.4	99.9	1290	3	US-09-993-059-3	Sequence 3, Appl
26	1252.4	99.9	1290	5	US-10-103-327-3	Sequence 3, Appl
27	1252.4	99.9	1290	7	US-10-602-219-3	Sequence 3, Appl
28	1252.4	99.9	1290	7	US-10-602-220-3	Sequence 3, Appl
29	1252.4	99.9	1290	7	US-10-411-037-67	Sequence 67, Appl
30	1252.4	99.9	1290	7	US-10-411-026-67	Sequence 67, Appl
31	1252.4	99.9	1290	7	US-10-410-962-67	Sequence 67, Appl
32	1252.4	99.9	1290	7	US-10-411-049-67	Sequence 67, Appl
33	1252.4	99.9	1290	7	US-10-410-930-67	Sequence 67, Appl
34	1252.4	99.9	1290	7	US-10-410-997-67	Sequence 67, Appl
35	1252.4	99.9	1290	7	US-10-411-012-67	Sequence 67, Appl
36	1252.4	99.9	1290	7	US-10-410-913-67	Sequence 67, Appl
37	1252.4	99.9	1290	8	US-10-851-388-3	Sequence 3, Appl
38	1252.4	99.9	1290	8	US-10-410-980-67	Sequence 67, Appl
39	1252.4	99.9	1290	9	US-10-410-987-67	Sequence 67, Appl
40	1252.4	99.9	1290	9	US-10-984-389-3	Sequence 3, Appl
41	1252.4	99.9	1296	3	US-09-993-059-9	Sequence 9, Appl
42	1252.4	99.9	1296	5	US-10-103-327-9	Sequence 9, Appl
43	1252.4	99.9	1296	7	US-10-602-219-9	Sequence 9, Appl
44	1252.4	99.9	1296	7	US-10-602-220-9	Sequence 9, Appl
45	1252.4	99.9	1296	8	US-10-851-388-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Query Match	100.0%	Score 1254;	DB 3;	Length 1254;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1254;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTCTGGCC	60	
Db	1	ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCTCTGGCC	60	
QY	61	CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGACGCT	120	
Db	61	CTCGTTTCTTGGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGACGCT	120	
QY	121	ACCATGGCTGGCTGGGACCTTTCATGTGCAACCTTGCATGCCAGGAGAGGCCA	180	
Db	121	ACCATGGCTGGCTGGGACCTTTCATGTGCAACCTTGCATGCCAGGAGAGGCCA	180	
QY	181	GATTCTTCATCAGTCAGAGAGCTTTCATTCAGAGAGGAGAGCTCATGTCTCAGAAGGC	240	
Db	181	GATTCTTCATCAGTCAGAGAGCTTTCATTCAGAGAGGAGAGCTCATGTCTCAGAAGGC	240	
QY	241	TGGAAGGATGCAAGTTATCAGTACCTCTGATGTGATGGTCTCCCAAGA	300	
Db	241	TGGAAGGATGCAAGTTATCAGTACCTCTGATGTGATGGTCTCCCAAGA	300	

Db 241 TGGAGGATGTCAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAAGA 300
Qy 301 GATTTCAGAGGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATCGGATTGCCAGCTA 360
Db 301 GATTTCAGAGGAGACTTCAGGCGAGACCTTCAGCGCTTTCCTCATCGGATTGCCAGCTA 360
Qy 361 GCTAAATTATGTTTCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420
Db 361 GCTAAATTATGTTTCACAGCAAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420
Qy 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATATCTACGACATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGACTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
Db 481 GACTGGGGAGTAGACTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
Qy 541 GCAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTGTGTAC 600
Db 541 GCAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCATACATCGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCATACATCGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAAGTATAAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGG 780
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCTTCAGCTGGAAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCTTCAGCTGGAAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGGACGTAATTCGCAATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCAAGCTCTCCTTCAGGATAAGGACGTAATTCGCAATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATGAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATGAACCGGACGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAAGAGTGGCTCTAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAAGAGTGGCTCTAATCCT 1140
Qy 1141 GCCTGCTTCATCAGACGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCAGACGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGACCTGTTTGTCTTCAAGCTATAA 1254

RESULT 2

US-10-103-327-15
; Sequence 15, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Query Match 100.0%; Score 1254; DB 5; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGCCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGCCTTCGCTTCCTGGCC 60
Qy 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGCTGCTGCTGCTGAGGAGCTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Db 121 ACCATGGCTGCTGCTGCTGAGGAGCTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 181 GATTCCTGTCATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Db 181 GATTCCTGTCATCAGTGAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
Qy 241 TGAAGGATGAGGTTATGAGTACCTCTGTCATGATGATGACTGTTGGATGCTCCCAAGA 300
Db 241 TGAAGGATGAGGTTATGAGTACCTCTGTCATGATGATGACTGTTGGATGCTCCCAAGA 300
Qy 301 GATTCAGAGGAGAGCTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAGAGGAGAGCTTCAGGCGAGACCTTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTCACAGCAAGGAGCTGAAGCTAGGGAATTTATGCAGATGTTGAAATAA 420
Db 361 GCTAATTTATGTTTCACAGCAAGGAGCTGAAGCTAGGGAATTTATGCAGATGTTGAAATAA 420
Qy 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACTAGCAATTCGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCGAGGCTTCCTGGGAGTTTGGATACTAGCAATTCGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGACTGCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
Db 481 GACTGGGGAGTAGACTGCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTG 540
Qy 541 GCAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTGTGTAC 600
Db 541 GCAGATGGTTTAAAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAGACATTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCATACATCGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAAGTATAAAG 720
Db 661 CAGTACTGCATACATCGGCGAAATTTTGTGACATTTGATGATGATTCCTGGAAAAGTATAAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGTCGACCGAGG 780

Db 721 AGTATCTTGACCTGGACATCTTTTAAACGAGAGAGAAATGTTGATGTTGTCGACGAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATGTCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATGTCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGCTACACGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACACGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCAGGCTTAGGCTGGCTGTAGCTATGATATAACCGGACGAGATGGT 1080
Db 1021 GAACGACCTCTCAGGCTTAGGCTGGCTGTAGCTATGATATAACCGGACGAGATGGT 1080
Qy 1081 GGACCTCGCTCTTATACCAATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACCAATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGA 1200
Db 1141 GCCTGCTTATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGA 1200
Qy 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 3

US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grilli, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15
Query Match 100.0%; Score 1254; DB 7; Length 1254;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTCGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTCGGCC 60
Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCGCAGGAGAGCCA 180
Qy 181 GATTCTCTGATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTCTCTGATCAGTCAGAAAGCTCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGGAAAGATGCAAGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
Db 241 TGGAAAGATGCAAGTTATGAGTACCTCTGCAATTTGATGACTGTTGGATGGCTCCCCAAAGA 300
Qy 301 GATTCAAGAGGACAGCTTCAGGACAGCCCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
Db 301 GATTCAAGAGGACAGCTTCAGGACAGCCCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 361 GCTAATTTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCAATGTTGGAAATAAA 420
Db 361 GCTAATTTATGTTTACAGCAAGAGCTGAAGCTAGGATTTATGCAATGTTGGAAATAAA 420
Qy 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGATTGGAAAATTTG 540
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGATTGGAAAATTTG 540
Qy 541 GCAGATGTTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTGTAC 600
Db 541 GCAGATGTTTATAAGCACATGCTTGGCCCTGTAATAGGACTGGCAGAGCAATTTGTGTAC 600
Qy 601 TCTGTGAGTGGCTCTTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Db 601 TCTGTGAGTGGCTCTTTTATATGTCGCCCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGCTGGGAAAATTTTGTGCAATTTGATGATTTCTTGGAAAAGTATAAG 720
Db 661 CAGTACTGCAATCAGCTGGGAAAATTTTGTGCAATTTGATGATTTCTTGGAAAAGTATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGCTGACACAGGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTTGGCAACTTTGGGCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
Qy 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATGTCATCAAT 960
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATGTCATCAAT 960

Qy 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGGCTTCAGCTATAA 1254

RESULT 5

US-10-851-388-15

; Sequence 15, Application US/10851388

; Publication No. US20040234516A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/851,388

; CURRENT FILING DATE: 2004-05-21

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1254)

US-10-851-388-15

Query Match 100.0%; Score 1254; DB 8; Length 1254;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACATACATCTGGGCTGCGGCTTGCGCTTCGCTTCCTGGCC 60
Qy 61 CTGCTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGAGCGCCT 120
Db 61 CTGCTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGAGCGCCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 181 GATTTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 181 GATTTCCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAAGA 300
Db 241 TGAAGAGTGCAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCCAAAGA 300
Qy 301 GATTTCAGAGGACGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Db 301 GATTTCAGAGGACGACTTCAGGAGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360
Qy 361 GCTAATTTATGTCACAGCAAGAGCTCAAGCTAGGAGTTATGACAGATGTTGGAATATAA 420
Db 361 GCTAATTTATGTCACAGCAAGAGCTCAAGCTAGGAGTTATGACAGATGTTGGAATATAA 420
Qy 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACGACATGATGCCAGAGCTTTGCT 480
Db 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGGATACGACATGATGCCAGAGCTTTGCT 480
Qy 481 GACTGGGGAGTAGTCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540

Db 481 GACTGGGGAGTAGTCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
Qy 541 GCAGATGGTTTAAAGCAATGCTGCTTGGCCCTGAAATAGGAGCTGCGACAGCAATGTTGTG 600
Db 541 GCAGATGGTTTAAAGCAATGCTGCTTGGCCCTGAAATAGGAGCTGCGACAGCAATGTTGTG 600
Qy 601 TCCTGTGAGTGGGCTCTTTATATGTTGTCCTTCAAAAGGCCAATTTATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGGCTCTTTATATGTTGTCCTTCAAAAGGCCAATTTATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCCAAATTTTGTGACATGATGATCTCTGAAAAGTATAAAG 720
Db 661 CAGTACTGCAATCACTGGCCAAATTTTGTGACATGATGATCTCTGAAAAGTATAAAG 720
Qy 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGTCGACCAAGG 780
Db 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGTCGACCAAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTGGCCCTCAGCTGGAATCAGCA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCAATTTGGCCCTCAGCTGGAATCAGCA 840
Qy 841 GTAACCTCAGATGCCCTCTGCGCTATCATGCTGCTCTTTTATTTTATGTTCTTAATGACCTC 900
Db 841 GTAACCTCAGATGCCCTCTGCGCTATCATGCTGCTCTTTTATTTTATGTTCTTAATGACCTC 900
Qy 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACATCATGAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
Qy 1021 GRACGACCTCTCTCAGGCTTAGGCTGCGCTGATGATGATAAACCGGAGGAGATGGT 1080
Db 1021 GRACGACCTCTCTCAGGCTTAGGCTGCGCTGATGATGATAAACCGGAGGAGATGGT 1080
Qy 1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Db 1081 GGAACCTCGCTCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAGTACGGTTCATGATGAGTACT 1200
Db 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGAAAAGGAGTACGGTTCATGATGAGTACT 1200
Qy 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTCAGCTATAA 1254

RESULT 6

US-10-984-389-15

; Sequence 15, Application US/10984389

; Publication No. US20050125859A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/984,389

; CURRENT FILING DATE: 2004-11-08

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

Db 361 GCTAAATATTGTTCAGCAAAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy 421 ACCTGCCAGGCTTCCCTGGGAGTTTGGATACACATGATGATGCCAGACCTTTGCT 480
Db 421 ACCTGCCAGGCTTCCCTGGGAGTTTGGATACACATGATGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGGAGTAGACTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
Db 481 GACTGGGGAGTAGACTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
Qy 541 GCAGATGTTATAGACATGCTGCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Db 541 GCAGATGTTATAGACATGCTGCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGGCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGGCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTCGAAGGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTCGAAGGATATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGATTTGATGTTGTCGACCGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGATTTGATGTTGTCGACCGAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACATGAGTGGCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 900
Db 841 GTAACATGAGTGGCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 900
Qy 901 CGACATATGAGCTTCAAGCCAAAGCTCTCTTCAGGATTAAGAGCTTAATGCCATCAAT 960
Db 901 CGACATATGAGCTTCAAGCCAAAGCTCTCTTCAGGATTAAGAGCTTAATGCCATCAAT 960
Qy 961 CAGGACCTTGGCAAGAGGTTACAGCTTACAGAGGAGACAACTTTGAAGTGTGG 1020
Db 961 CAGGACCTTGGCAAGAGGTTACAGCTTACAGAGGAGACAACTTTGAAGTGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAACCGGACGAGATTGGT 1080
Qy 1081 GGACCTGCTCTTATACATGCGAGTTGCTTCCCTGGGTTAAGAGTGGCTGTATCCT 1140
Db 1081 GGACCTGCTCTTATACATGCGAGTTGCTTCCCTGGGTTAAGAGTGGCTGTATCCT 1140
Qy 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Db 1141 GCCTGCTTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGAGCT 1200
Qy 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254

RESULT 8

US-10-103-327-11
; Sequence 11, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10103,327
; CURRENT FILING DATE: 2002-03-29

; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11

Query Match 99.9%; Score 1252.4; DB 5; Length 1266;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCCCTTCCTGCGCC 60
Db 1 ATCAGCTGAGGAAACCCAGAACTACATCTGGGCTGCGCTTGGCTTCCCTTCCTGCGCC 60
Qy 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGAGAGAGGCA 180
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGAATGAGAGAGGCA 180
Qy 181 GATTCTGTCATCAGTCAGAGCTCTTCATGGAGAGCTCTTCATGGAGAGCTCATGGTCTCAGAGGC 240
Db 181 GATTCTGTCATCAGTCAGAGCTCTTCATGGAGAGCTCTTCATGGAGAGCTCATGGTCTCAGAGGC 240
Qy 241 TGAAGAGTGCAGCTTATGATACCTCTGCAATGATGACCTGTTGGATGGCTTCCCAAGA 300
Db 241 TGAAGAGTGCAGCTTATGATACCTCTGCAATGATGACCTGTTGGATGGCTTCCCAAGA 300
Qy 301 GATTCAAGAGGAGACTTTCAGGAGAGCTTTCAGGAGAGCTTTCATGGATGGCTTGGCAAGT 360
Db 301 GATTCAAGAGGAGACTTTCAGGAGAGCTTTCAGGAGAGCTTTCATGGATGGCTTGGCAAGT 360
Qy 361 GCTAATTTGTTTTCAGAGAGGAGCTTTCAGGAGAGCTTTCATGGATGGCTTGGCAAGT 420
Db 361 GCTAATTTGTTTTCAGAGAGGAGCTTTCAGGAGAGCTTTCATGGATGGCTTGGCAAGT 420
Qy 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGATGATGATGATGAT 480
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGATGATGATGATGAT 480
Qy 481 GACTGGGGAGTAGACTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
Db 481 GACTGGGGAGTAGACTGCTAAATAATTTGATGTTGTTTACTGTGACAGTTTGGAAAAATTG 540
Qy 541 GCAGATGTTATAGACATGCTGCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Db 541 GCAGATGTTATAGACATGCTGCTGGCCCTGAATAGGACTGCGAAGCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGGCCAATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGGCCAATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTCGAAGGATATAAG 720
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTTCTCGAAGGATATAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGATTTGATGTTGTCGACCGAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGATTTGATGTTGTCGACCGAGG 780
Qy 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840

1021	GAAGGACCTCTCTCAGCGCTTAGCCTGGGCTGTAGCTATGATATAAACCGCAGAGATTGGT	1080
1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGATGGCCTGTAAATCCT	1140
1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGATGGCCTGTAAATCCT	1140
1141	GCCTGCTTCAATCACACAGCTCCTCCCTGTGAAAGGAAAGCTAGGGTTCTATGAATGGACT	1200
1141	GCCTGCTTCAATCACACAGCTCCTCCCTGTGAAAGGAAAGCTAGGGTTCTATGAATGGACT	1200
1201	TCAAGGTTAAGAGTGCATATAATCCACAGGCACCTGTTTGTCTTCAAGCTATATA	1254
1201	TCAAGGTTAAGAGTGCATATAATCCACAGGCACCTGTTTGTCTTCAAGCTAGAA	1254

RESULT 10

US-10-602-220-11

Sequence 11, Application US/10602220
Publication No. US20040023281A1

GENERAL INFORMATION:
APPLICANT: Large Scale Biology Corporation
APPLICANT: Turpen, Thomas H.
APPLICANT: Kumagai, Monto H.
APPLICANT: Pogue, Gregory P.
APPLICANT: Erwin, Robert L.
APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSBC-0087-CP07B
CURRENT APPLICATION NUMBER: US/10/602,220
CURRENT FILING DATE: 2003-06-23
PRIOR APPLICATION NUMBER: 09/993,059
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: 09/626,127
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 09/316,572
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: 08/324,003
PRIOR FILING DATE: 1994-10-14
PRIOR APPLICATION NUMBER: 08/176,414
PRIOR FILING DATE: 1993-12-29
PRIOR APPLICATION NUMBER: 07/997,733
PRIOR FILING DATE: 1992-12-30
PRIOR APPLICATION NUMBER: 08/184,237
PRIOR FILING DATE: 1994-01-19
PRIOR APPLICATION NUMBER: 07/923,692
PRIOR FILING DATE: 1992-07-31
PRIOR APPLICATION NUMBER: 07/600,244
PRIOR FILING DATE: 1990-10-22
PRIOR APPLICATION NUMBER: 07/641,617
PRIOR FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 11
LENGTH: 1266
TYPE: DNA
ORGANISM: Homo sapiens
US-10-602-220-11

	Query Match	99.9%;	Score 1252.4;	DB 7;	Length 1266;
	Best Local Similarity	99.9%;	Pred. No. 0;		
	Matches 1253;	Conservative	0;	Mismatches	1;
				Indels	0;
				Gaps	0;
Qy	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCGCTTCGCTTCCTGGCC	60		
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCGCTTCGCTTCCTGGCC	60		
Qy	61	CTCGTTTCTTGGACATCCCTGGGGCTAGAGCACTGGACAATGATTTGGCAAGACGCCT	120		
Db	61	CTCGTTTCTTGGACATCCCTGGGGCTAGAGCACTGGACAATGATTTGGCAAGACGCCT	120		
Qy	121	ACCATGGCTGGCTGCATCTGGGAGCGCTTCATCTGCACCTTGACCTGCCAGGAAGACCA	180		

601	TCCTGTGAGTGGCCCTCTTTATATGTGTGGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA	660
661	CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTCATGATTCCTCGAAAGCTATAAAG	720
661	CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTCATGATTCCTCGAAAGCTATAAAG	720
721	AGTATCTTTGGACTGGACATCTTTTAAACGAGAGAGAAATGTTGATGTTGTGACACAGGG	780
721	AGTATCTTTGGACTGGACATCTTTTAAACGAGAGAGAAATGTTGATGTTGTGACACAGGG	780
781	GGTTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTTCAGCTGGAATCAGCAA	840
781	GGTTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCTTCAGCTGGAATCAGCAA	840
841	GTAACCTCAGATGGCCCTCTGGGCTTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC	900
841	GTAACCTCAGATGGCCCTCTGGGCTTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC	900
901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCGATCAAT	960
901	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTCGATCAAT	960
961	CAGGACCCCTTTGGGCAAGAGGCTTACGAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
961	CAGGACCCCTTTGGGCAAGAGGCTTACGAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
1021	GAACACCTCTCTCAGGCTTAGCCCTTAGGCTGCTAGCTATGATATAACCGGACAGGATTTGGT	1080
1021	GAACACCTCTCTCAGGCTTAGCCCTTAGGCTGCTAGCTATGATATAACCGGACAGGATTTGGT	1080
1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT	1140
1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTAAATCCT	1140
1141	GCCTGTTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGCTTCTATGAATGAGCT	1200
1141	GCCTGTTTCATCACACAGCTCTCCCTGTGAAAGGAGCTAGGCTTCTATGAATGAGCT	1200
1201	TCAAGGTTAAGAGTCACTAAATCCCAAGGACACTGTTTTGCTTCAGCTATAA	1254
1201	TCAAGGTTAAGAGTCACTAAATCCCAAGGACACTGTTTTGCTTCAGCTATAA	1254
RESULT 12		
US-10-984-389-11		
; Sequence 11, Application US/10984389		
; Publication No. US20050125859A1		
; GENERAL INFORMATION:		
; APPLICANT: GARGER, Stephen A.		
; APPLICANT: TURPEN, Thomas H.		
; APPLICANT: KUMAGAI, Monto H.		
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN		
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION		
; FILE REFERENCE: 008010087CPUS06		
; CURRENT APPLICATION NUMBER: US/10/984,389		
; PRIOR FILING DATE: 2004-11-08		
; PRIOR APPLICATION NUMBER: US/09/993,059		
; PRIOR FILING DATE: 2001-11-13		
; NUMBER OF SEQ ID NOS: 37		
; SOFTWARE: FastSeq for Windows Version 4.0		
; SEQ ID NO 11		
; LENGTH: 1266		
; TYPE: DNA		
; ORGANISM: Homo sapiens		
; FEATURE:		
; NAME/KEY: CDS		
; LOCATION: (1)...(1266)		
US-10-984-389-11		
Query Match 99.9%; Score 1252.4; DB 9; Length 1266;		
Best Local Similarity 99.9%; Pred. No. 0;		
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	1	ATCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
DB	1	ATCAGCTCAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY	61	CTGCTTTCTCTGGACATCTCCCTGGGCTAGAGCACTGGCAATGATGGCAAGACGCCT 120
DB	61	CTGCTTTCTCTGGACATCTCCCTGGGCTAGAGCACTGGCAATGATGGCAAGACGCCT 120
QY	121	ACCATGGCTGGTGCATCTGGGAGCGCTTCATCTGCAACCTTGACTGCAGAGAGGCCA 180
DB	121	ACCATGGCTGGTGCATCTGGGAGCGCTTCATCTGCAACCTTGACTGCAGAGAGGCCA 180
QY	181	GATTCCTGCATCAGTCAGAGAGCTTCTCATGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
DB	181	GATTCCTGCATCAGTCAGAGAGCTTCTCATGAGATGGCAGAGCTCATGCTCTCAGAGGC 240
QY	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
DB	241	TGGAAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCAAGA 300
QY	301	GATTCAGAGGCGAGCTTCAGGCGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
DB	301	GATTCAGAGGCGAGCTTCAGGCGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY	361	GCTAAATTATCTTCACAGCAAGAGCTGAAGCTAGGATTTATGACAGATGTTGGAAATAA 420
DB	361	GCTAAATTATCTTCACAGCAAGAGCTGAAGCTAGGATTTATGACAGATGTTGGAAATAA 420
QY	421	ACCTGGCGAGGCTTCCCTGGGATTTTGGATACACGATTTGATGATGATGATGATGATGAT 480
DB	421	ACCTGGCGAGGCTTCCCTGGGATTTTGGATACACGATTTGATGATGATGATGATGATGAT 480
QY	481	GATCTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
DB	481	GATCTGGGAGTAGATCTGCTAAATTTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
QY	541	GCAGATGTTTATAGCACATCTCTTGGCCCTGTAATAGGACTGGCAGAGGATTTGTGTAC 600
DB	541	GCAGATGTTTATAGCACATCTCTTGGCCCTGTAATAGGACTGGCAGAGGATTTGTGTAC 600
QY	601	TCCTGTGAGTGGCCCTCTTTATATGTGTGGCCCTTTCAAAGGCCCAATATACAGAAATCCGA 660

Qy 961 CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAATTTGAAGTGG 1020
Db CAGGACCCCTTGGGCAAGGCTACAGCTTAGACAGGAGACAATTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATAACCGCAGGAGATTGGT 1080
Qy 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Db 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCTCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
Db 1141 GCCTGCTTCATCACAGCTCTCTCCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200
Qy 1201 TCAAGGTTAAGAGTCATATAATCCACAGGCACTGTTTGTCTCAGCTATTA 1254
Db 1201 TCAAGGTTAAGAGTCATATAATCCACAGGCACTGTTTGTCTCAGCTATTA 1254

RESULT 15

US-10-602-219-7
; Sequence 7, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: USBC-0087-CP09B
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: US/10/602,219
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-602-219-7

Query Match 99.98; Score 1252.4; DB 7; Length 1278;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCGTGGCC 60
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCGTGGCC 60

Qy 61 CTGTTTCTGGGACATCCCTGGGCTAGAGCAATGGATTTGGCAAGAGACGCT 120
Db 61 CTGTTTCTGGGACATCCCTGGGCTAGAGCAATGGATTTGGCAAGAGACGCT 120
Qy 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Db 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
Qy 181 GATTTCCTGCACTAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
Db 181 GATTTCCTGCACTAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
Qy 241 TGAAGGATGCAGGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 241 TGAAGGATGCAGGTTATGATGACTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Qy 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTCTCATGGATTTGCCAGCTA 360
Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTCTCATGGATTTGCCAGCTA 360
Qy 361 GCTAATTTATGTTTCAAGCAAGGACTGAAGCTAGGGATTTATGCAGATTTGGAAATATAA 420
Db 361 GCTAATTTATGTTTCAAGCAAGGACTGAAGCTAGGGATTTATGCAGATTTGGAAATATAA 420
Qy 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACACTGACATTTGATGCCAGACCTTTGCT 480
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACACTGACATTTGATGCCAGACCTTTGCT 480
Qy 481 GACTGGGAGTACATCTGCTTAAATTTGATGTTTGTACGTGACACGTTTGGAAATTTG 540
Db 481 GACTGGGAGTACATCTGCTTAAATTTGATGTTTGTACGTGACACGTTTGGAAATTTG 540
Qy 541 GCAGATGGTTTATAGCACATCTCTTGGCCCTGAATAGGACTGGCAGAACCATTTGTGTAC 600
Db 541 GCAGATGGTTTATAGCACATCTCTTGGCCCTGAATAGGACTGGCAGAACCATTTGTGTAC 600
Qy 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTCAAAGGCCAATATATACAGAAATCCGA 660
Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTCAAAGGCCAATATATACAGAAATCCGA 660
Qy 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATTTGATGATTTCTGGAAGATTAAG 720
Db 661 CAGTACTGCAATCAGCTGGGAAATTTTGTGACATTTGATGATTTCTGGAAGATTAAG 720
Qy 721 AGTATCTTGGACTGGACATCTTTTAAACCGAGAGAGAAATTTGATTTGTCGACCAAGG 780
Db 721 AGTATCTTGGACTGGACATCTTTTAAACCGAGAGAGAAATTTGATTTGTCGACCAAGG 780
Qy 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Db 781 GGTGGAATGACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 841 GTAACTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTTATGCTTAATGACCTC 900
Db 841 GTAACTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTTATTTTATGCTTAATGACCTC 900
Qy 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 961 CAGGACCCCTTGGGCAAGCAAGGTTAGACAGGAGACAACTTTGAAGTGG 1020
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTAGACAGGAGACAACTTTGAAGTGG 1020
Qy 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATAACCGCAGGAGATTGGT 1080
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTAGCTATGATAACCGCAGGAGATTGGT 1080
Qy 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTTGAATCCT 1140
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTTGAATCCT 1140
Qy 1141 GCCTGCTTCATCACAGCTCTCTCCTCTGTGAAAAGGAGCTAGGGTTCTATGAATGGACT 1200

Db 1141 GCTGCTTCATCACAGCTCCCTCTGTAAGAGGAGCTAGGGTTCTATGATGGACT 1200
QY 1201 TCAAGGTTAAGAAGTCAACATAATCCACAGGCACTGTTTTGCTTCAGCTATAA 1254
Db 1201 TCAAGGTTAAGAAGTCAACATAATCCACAGGCACTGTTTTGCTTCAGCTAGAA 1254

Search completed: December 31, 2005, 23:36:16
Job time : 1165 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 31, 2005, 21:00:36 ; Search time 247 Seconds
(without alignments)
2661.468 Million cell updates/sec

Title: US-10-602-220-15
Perfect score: 1254
Sequence: 1 atgcagctgagggaaccaga.....ctgtttgttcagctataa 1254

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New.*
1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1252.4	99.9	1355	6	US-10-981-267-26
2	1252.4	99.9	1547	6	US-10-981-267-25
3	37	3.0	2779	6	US-10-750-185-40681
4	36.6	2.9	1306	6	US-10-750-185-42255
5	35	2.8	321019	6	US-10-995-561-13204
6	34	2.7	1932	6	US-10-750-185-55363
7	33	2.6	1437	6	US-10-750-185-43877
8	32.8	2.6	1666	6	US-10-821-234-149
9	32.8	2.6	146658	7	US-11-121-086-68
10	32.8	2.6	188682	7	US-11-112-908-23
11	32.8	2.6	190882	7	US-11-121-086-69
12	32.4	2.6	1685	6	US-10-750-185-53897
13	32.2	2.6	86081	6	US-10-995-561-13246
14	32	2.6	1254	6	US-10-750-185-29504
15	31.8	2.5	201	6	US-10-995-561-29100
16	31.8	2.5	4847	6	US-10-750-185-60280
17	31.6	2.5	2787	6	US-10-841-129-3
18	31.6	2.5	1125000	6	US-10-995-561-13286
19	31.4	2.5	119160	7	US-11-121-086-12
20	31.4	2.5	161874	7	US-11-121-086-75
21	31.2	2.5	1522	6	US-10-750-185-37196
22	31	2.5	854	6	US-10-750-185-28311
23	31	2.5	1084	6	US-10-750-185-61883

Sequence 37071, A
Sequence 3, Appli
Sequence 6485, A
Sequence 121, App
Sequence 357, App
Sequence 54302, A
Sequence 13493, A
Sequence 42925, A
Sequence 66, Appl
Sequence 60, Appl
Sequence 39400, A
Sequence 1, Appli
Sequence 13227, A
Sequence 43163, A
Sequence 28504, A
Sequence 39271, A
Sequence 35062, A
Sequence 63028, A
Sequence 4, Appli
Sequence 5, Appli
Sequence 2, Appli
Sequence 6, Appli

US-10-750-185-37071
US-11-121-086-3
US-10-750-185-64685
US-10-131-826A-121
US-10-821-234-357
US-10-750-185-54302
US-10-995-561-13493
US-10-750-185-42925
US-11-121-086-66
US-11-121-086-60
US-10-750-185-39400
US-11-121-086-1
US-10-995-561-13227
US-10-750-185-43163
US-10-750-185-28504
US-10-750-185-39271
US-10-750-185-35062
US-10-750-185-63028
US-10-509-921-4
US-10-509-921-5
US-10-509-921-2
US-10-509-921-6

ALIGNMENTS

RESULT 1

US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-981-267-26

Query Match 99.9%; Score 1252.4; DB 6; Length 1355;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCAGCTGAGGAGCCAGCACTACATCTGGGCTGGCGCTTCGCTTCCTCTCGGCC 60
Db 48 ATGCAGCTGAGGAGCCAGCACTACATCTGGGCTGGCGCTTCGCTTCCTCTCGGCC 107
QY 61 CTCGTTTCTGGGACATCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCT 120
Db 108 CTCGTTTCTGGGACATCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCCT 167

121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
168 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 227
181 GATTCTGTCATAGTGAAGAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
228 GATTCTGTCATAGTGAAGAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 287
241 TGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
288 TGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 347
301 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGAGATGGCAGAGCTA 360
348 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGAGATGGCAGAGCTA 407
361 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTAGGAGTTATCGAGATGTTGGAATAAA 420
408 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTAGGAGTTATCGAGATGTTGGAATAAA 467
421 ACCTGGCAGGCTTCCCTGGGAGCTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
468 ACCTGGCAGGCTTCCCTGGGAGCTTTGGATACCTACGACATTCATGCCAGACCTTTGCT 527
481 GACTGGGAGCTAGATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTG 540
528 GACTGGGAGTGAATCTGCTAAATTTGATGGTGTGTTACTGTGACAGTTTGGAAAATTG 587
541 GCAGATGGTTATGAAGCAGATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATGTTGTAC 600
588 GCAGATGGTTATGAAGCAGATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATGTTGTAC 647
601 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 660
648 TCCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA 707
661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 720
708 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATTCCTGGAAAAGTATAAAG 767
721 AGTATCTTGACATGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTGACCAAGG 780
768 AGTATCTTGACATGGACATCTTTTAAACAGGAGAGAAATTTGATGTTGTGACCAAGG 827
781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
828 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 887
841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 900
888 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTTCTAATGACCTC 947
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGACGTAATGCAATCAAT 960
948 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTTCAAGGATAAGGACGTAATGCAATCAAT 1007
961 CAGAACCCCTTGGCAGAGAGGGTACACGCTTAGACAGGAGACAACTTTGAAGTGG 1020
1008 CAGAACCCCTTGGCAGAGAGGGTACACGCTTAGACAGGAGACAACTTTGAAGTGG 1067
1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATTAACCCGAGGAGATGGT 1080
1068 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATTAACCCGAGGAGATGGT 1127
1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
1128 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1187
1141 GCCTGCTTATCAGAGCTCTCCCTGTGAAAGAGAGCTAGGGTCTATGATGAGCT 1200
1188 GCCTGCTTATCAGAGCTCTCCCTGTGAAAGAGAGCTAGGGTCTATGATGAGCT 1247
1201 TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTATAA 1254

1248 TCAAGGTTAAGAGTCAATAAATCCCAAGGCACTGTTTTGCTTCAGCTAGAA 1301
RESULT 2
US-10-981-267-25
; Sequence 25, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence
US-10-981-267-25

Query Match 99.9%; Score 1252.4; DB 6; Length 1547;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTCGCC 60
Db 51 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCTCGCC 110
Qy 61 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 120
Db 111 CTCGTTTCTGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 170
Qy 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Db 171 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 230
Qy 181 GATTCTTCATCAGTGAAGAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Db 231 GATTCTTCATCAGTGAAGAGCTTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 290
Qy 241 TGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 300
Db 291 TGAAGGATGCAAGGTTATGAGTACCTCTGCAATGATGACTGTTGGATGGCTCCCAAGA 350
Qy 301 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGAGATGGCAGACTA 360
Db 351 GATTCAAGAGCAGACTTCAGGCAGACCTTCAGCGCTTCTCATGGAGATGGCAGACTA 410
Qy 361 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATAAA 420
Db 411 GCTAATTTATGTTACAGCAAGAGGACTGAAGCTAGGAGTTTATGAGATGTTGGAATAAA 470
Qy 421 ACCTGGCAGGCTTCCCTGGGAGTTCATGAGATGTTTGGATACCTACGACATTCATGATGCCAGACCTTTGCT 480

Db	471	ACCTGCGAGGCTTCCCTGGGAGTTTGGATACACGACATGTGAATGCCAGACCTTTGCT	530
Qy	481	GACTGGGGAGTAGACTCTGTAATAATTTGATGGTTTACTGTGACAGTTTGGAAAAATTG	540
Db	531	GACTGGGGAGTAGACTCTGTAATAATTTGATGGTTTACTGTGACAGTTTGGAAAAATTG	590
Qy	541	GCAGATGGTTATAAGCAATGTCCTTGGCCCTGAAATAGGACTGCGAGAAGCATTTGTGTAC	600
Db	591	GCAGATGGTTATAAGCAATGTCCTTGGCCCTGAAATAGGACTGCGAGAAGCATTTGTGTAC	650
Qy	601	TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA	660
Db	651	TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA	710
Qy	661	CAGTACTGCAATCACTGCGCAATTTTGTCTGACATTCGATTCCTCGAAAAGTATAAAG	720
Db	711	CAGTACTGCAATCACTGCGCAATTTTGTCTGACATTCGATTCCTCGAAAAGTATAAAG	770
Qy	721	AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTGATTTGTCTGGACCAAGG	780
Db	771	AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTGATTTGTCTGGACCAAGG	830
Qy	781	GGTTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCA	840
Db	831	GGTTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCA	890
Qy	841	GTAACCTCAGATGGCCCTCTGGCTCATCATGGCTGCTCTTTATTTCATGTCCTTAATGACCTC	900
Db	891	GTAACCTCAGATGGCCCTCTGGCTCATCATGGCTGCTCTTTATTTCATGTCCTTAATGACCTC	950
Qy	901	CGACATCATGACCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGACGTAAATTGCCATCAAT	960
Db	951	CGACATCATGACCCCTCAAGCCAAAGCTCTCTCTTCAGGATAAGGACGTAAATTGCCATCAAT	1010
Qy	961	CAGSACCCCTTGGCAAGCAAGGATACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
Db	1011	CAGSACCCCTTGGCAAGCAAGGATACAGCTTAGACAGGAGACAACTTTGAAGTGTGG	1070
Qy	1021	GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCCGGCAGGAGATTGGT	1080
Db	1071	GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATAAACCCGGCAGGAGATTGGT	1130
Qy	1081	GGACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAACTCT	1140
Db	1131	GGACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTAACTCT	1190
Qy	1141	GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
Db	1191	GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT	1250
Qy	1201	TCAAGGTTAAAGATCATATAATCCCAACAGGCACTCTTTTGTCTTACGCTATAA	1254
Db	1251	TCAAGGTTAAAGATCATATAATCCCAACAGGCACTCTTTTGTCTTACGCTATAA	1304

RESULT. T 3

```

US-10-750-185-40691/c
; Sequence 40691, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITING
; FILE REFERENCE: WM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,488

```

```

; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40681
; LENGTH: 2779
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-40681

```

RESULT 4

```

RES001.4
US-10/750-185-42255
; Sequence 42255, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42255
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bovine
US-10/750-185-42255

```

RESULT 5

US-10-995-561-13204/c
; Sequence 13204, Application US/10995561
; Publication No. US20050272054A1

```
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13204
; LENGTH: 321019
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(321019)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13204

Query Match      2.8%; Score 35; DB 6; Length 321019;
Best Local Similarity 49.1%; Pred. No. 11;
Matches 86; Conservative 2; Mismatches 87; Indels 0; Gaps 0;

QY      915 TCAGGCAAGAGCTCTCCTTCAGGATAGGAGCGTAATTGCCATCAATCAGGACCCCTTGGG 974
Db      301377 TCACAGAGGCTGACCTAGCAGACACRAAAGATTGTGTAAGCAGAGATGTCTAAAG 301318

QY      975 CAAGCAAGGGTACCAGCTTAGACAGGAGAGCAACTTTTGAAGTGTGGGAACGACTCTCTC 1034
Db      301317 CATGAGCTGGCTGCCTTTGTGAGGCGAGTGACATGAGTCCCTGGAAGAAACCTCTT 301258

QY      1035 AGGCTTAGCTGGGCTGTAGTATGATAAACCGCAGGAGATGGTGGACCTCGC 1089
Db      301257 AGCAGAAGCTCAACAGCAGCTGTCAAAGATCTGCAGGTGCACCTGGGGACTGGGC 301203

RESULT 6
US-10-750-185-55363/c
; Sequence 55363, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2003-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55363
; LENGTH: 1932
; TYPE: DNA
; ORGANISM: Bovine 19866880928582
US-10-750-185-55363

Query Match      2.7%; Score 34; DB 6; Length 1932;
Best Local Similarity 57.5%; Pred. No. 0.7;
Matches 61; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY      655 ATCCGACAGTACTGCATCACTGCGGAATTTTCTGCATGATGATGATCTCTGGAAGT 714
Db      1081 ATCTGAGATTTCTGAGATGATTTTCCAAAATTTATTTGGAGAAAGGATTTCTGAAGATT 1022

QY      715 ATAAAGAGTATCTTGGAGCTGGACATCTTTTAAACCAGGAGAGAAATTG 760
```

```
Db      1021 AGTGAAATTTTCTTGATGATACATACTTTTTTATCGACATTATATTG 976

RESULT 7
US-10-750-185-43877/c
; Sequence 43877, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43877
; LENGTH: 1437
; TYPE: DNA
; ORGANISM: Bovine 19866880874544
US-10-750-185-43877

Query Match      2.6%; Score 33; DB 6; Length 1437;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 69; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY      835 CACCAAGTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCTATGCTTAAT 894
Db      458 CAGGAATTTATTGGGTGGCTGACTGGGATATCCAGGAGAGAGCTGGCCTCAGGAATTCA 399

QY      895 GACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTGCC 954
Db      398 GAGCCAGAGCCCTTCTGGCTCCCTCTCATCTCTCCCGTGGTACTAGATGGTCTTTTGA 339

QY      955 ATCAATCAG 963
Db      338 ATGGAACAG 330

RESULT 8
US-10-821-234-149
; Sequence 149, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 149
; LENGTH: 1666
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-149

Query Match      2.6%; Score 32.8; DB 6; Length 1666;
Best Local Similarity 64.5%; Pred. No. 1.6;
Matches 49; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
```


Qy	127	GGTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCAGATTCC	186
Db	849	GGCTTCTGGTATGGCTCAGTTTTATGAGCCACGTTTAAAGGAGAGGAAACCAACATTTT	908
Qy	187	TGCATCAGTGAAGAAGCTCTTCATGAGATG	216
Db	909	TGTTTCCTTGAGTAACTCGTTATAGATTG	938

```

RESULT 13
US-10-995-561-13246/c
; Sequence 13246, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13246
; LENGTH: 86081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13246

```

	Query Match	2.6%	Score 32.2;	DB 6;	Length 86081;
	Best Local Similarity	48.0%;	Pred. No. 41;		
	Matches 82;	Conservative	3;	Mismatches 86;	Indels 0; Gaps 0;
Qy	104	GATTGGCAAGGAGCGCCTACCATGGGTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTG	163		
Db	68847	GGTGGGCAAGGCTTTGACCTTCCTCGCGGTAGTGGGTGCAGAYGTAGGTACCTG	68788		
Qy	164	ACTGCCAGGAAGAGCCAGATTCTCTGATCAGTGAGAAGCTTCTTCATGGAGATGGCAGAC	223		
Db	68787	CGTCTGGGTGGACTGAGGGCAGCATCAGCATGTGTCTCCAGGCGGCTGTCACTG	68728		
Qy	224	TCATGGTCTCAGAAGCGCTGGAAGGATGCGAGTTATCAGTACCTCTGCATTG	274		
Db	68727	GCAGGTGCCATCTCAGCTGVAATGCACTAGCACTGAGGGCCTGCGCCCTG	68577		

```

RESULT 14
US-10-750-185-29504
; Sequence 29504, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29504
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-29504

```

	Query Match	2.6%	Score 32;	DB 6;	Length 1254;
	Best Local Similarity	46.4%;	Pred. No. 2.6;		
	Matches 104;	Conservative 0;	Mismatches 120;	Indels 0;	Gaps 0;
Qy	215	TGGCAGAGCTCATGGTCTCAGAAAGGCTGGAAAGATGCAAGGTATGAGTACCTCTGCAATTG	274		
Db	440	TGGAGAAATCCATGGACAGAGGAGCTGGCAGGCTACAGTCCATGGGTGCTAAAGAAATTG	499		
Qy	275	ATGACTGTTGGATGGCTCCCAAAGATTTCAGAAGGCGAGACTTCAGGCGAGACCCCTCAGC	334		
Db	500	GACACAATTTAGTGAATAAACAATGATGCGAGAAAAGCCATAATACATTTTGGAGAGAAGT	559		
Qy	335	GCITTCCTCATGGATTCGCCAGCTACTAATTATGTTTCACAGCAAAAGGACTGAAGGCTAG	394		
Db	560	GGTCAGAGACGAGGACGATAAANAATGTTTAAATATATTCATGGCAACATCTCCGAGCAAA	619		
Qy	395	GGATTTTATGAGATGTTGGAAATAAAACCTGCGCAGGCTTCCCT	438		
Db	620	TTTTAAATAAATAAAAGTAAATAGACTCATTTCTGCTACTCT	663		

```

RESULT 15
US-10-995-561-29100/c
; Sequence 29100, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29100
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-29100

```

	Query Match	2.5%	Score 31.8;	DB 6;	Length 201;
	Best Local Similarity	49.1%;	Pred. No. 0.85;		
	Matches 84;	Conservative 0;	Mismatches 87;	Indels 0;	Gaps 0
Qy	104	GAITGGCAAGACGCCTACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTG	163		
Db	172	GGTGGGCAAGGCTTTTGACCTTGGCCGGTTAGTGGCGGTGCAGACGTAGGTACCTG	113		
Qy	164	ACTGCCAGGAAGAGCCAGATTCCTGCGCATCAGTGGAGAAGCTCTTCATGGAGATGGCAGAGC	223		
Db	112	CGTCTGGGGTGGACTGAGGGCAGCATCAGCATGTTGTTCTCCAGGGCGCTGTCCAGTGTG	53		
Qy	224	TCATSGTCTCAGAAGCTGGGAAGATGCAGGTTATAGTACCTCTGCAATTG	274		
Db	52	GCAAGGCTGCATCCAGCTGCAAAATGCACTAGCACTTGAGGGCCCTGGCCCTTG	2		

Search completed: December 31, 2005, 23:40:30
Job time : 252 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:04 ; Search time 410.772 Seconds
(without alignments)
2490.874 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140557_5421/app_query.fasta_1.782
-DB=GenEmbl -QWTF=fastcap -SURFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pt0 -NORM=ext -HRAPIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220 @CEN 1.1 5556 @runat_30122005_140557_5421 -NCPU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb.ba:*
2: gb.in:*
3: gb.env:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pr:*
9: gb.ro:*
10: gb.sts:*
11: gb.sy:*
12: gb.un:*
13: gb.vi:*
14: gb.ntg:*
15: gb.pli:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	89	100.0	210	6	AR101985 Sequence
2	89	100.0	210	6	BD064230 Therapy f
3	89	100.0	210	6	BD247363 Treatment

	4	89	100.0	210	6	CS105028	Sequence
5	89	100.0	210	6	AR234437	Sequence	
6	89	100.0	210	6	AR322463	Sequence	
7	89	100.0	210	6	AR372139	Sequence	
8	89	100.0	227	6	CO684109	Sequence	
9	89	100.0	270	6	CO685605	Sequence	
10	89	100.0	297	6	I30314	Sequence 11	
11	89	100.0	297	6	AX030666	Sequence	
12	89	100.0	1158	6	AR143996	Sequence	
13	89	100.0	1161	6	AR143995	Sequence	
14	89	100.0	1164	6	AR143994	Sequence	
15	89	100.0	1167	6	AR143993	Sequence	
16	89	100.0	1170	6	AR143992	Sequence	
17	89	100.0	1173	6	AR143991	Sequence	
18	89	100.0	1176	6	AR143990	Sequence	
19	89	100.0	1179	6	AR143989	Sequence	
20	89	100.0	1182	6	AR143988	Sequence	
21	89	100.0	1188	6	AR143987	Sequence	
22	89	100.0	1194	6	AR143986	Sequence	
23	89	100.0	1197	6	AR101990	Sequence	
24	89	100.0	1197	6	BD064235	Therapy f	
25	89	100.0	1197	6	BD247366	Treatment	
26	89	100.0	1197	6	CS105034	Sequence	
27	89	100.0	1197	6	AR234440	Sequence	
28	89	100.0	1197	6	AR322468	Sequence	
29	89	100.0	1197	6	AR372144	Sequence	
30	89	100.0	1214	6	AR380456	Sequence	
31	89	100.0	1214	8	HUMAGALA	Sequence	
32	89	100.0	1214	8	HUMAGALA	M13571 Human matur	
33	89	100.0	1254	6	AR653698	Sequence	
34	89	100.0	1254	6	AR654966	Sequence	
35	89	100.0	1266	6	AR653696	Sequence	
36	89	100.0	1266	6	AR654964	Sequence	
37	89	100.0	1272	6	AR653699	Sequence	
38	89	100.0	1272	6	AR654967	Sequence	
39	89	100.0	1278	6	AR653694	Sequence	
40	89	100.0	1278	6	AR654962	Sequence	
41	89	100.0	1284	6	AR653697	Sequence	
42	89	100.0	1284	6	AR654965	Sequence	
43	89	100.0	1290	6	CS135670	Sequence	
44	89	100.0	1290	6	AR653692	Sequence	
45	89	100.0	1290	6	AR654960	Sequence	

ALIGNMENTS

RESULT 1
AR101985
LOCUS AR101985 210 bp DNA linear PAT 14-FEB-2001
DEFINITION Sequence 19 from patent US 6083725.
ACCESSION AR101985
VERSION AR101985.1 GI:12812783
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.P., Borowski,M., Gillispie,P.P., Kinoshita,C.M.,
Tresco,D.A. and Williams,M.D.
TITLE Transfected human cells expressing human .alpha.-galactosidase A
protein
JOURNAL Patent: US 6083725-A 19 04-JUL-2000;
FEATURES Location/Qualifiers
source 1..210
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x AR101985 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAGTACATAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 2
BD064230
LOCUS BD064230 210 bp DNA linear PAT 27-AUG-2002
DEFINITION Therapy for alpha-galactosidase A deficiency.
ACCESSION BD064230
VERSION BD064230.1 GI:22609833
KEYWORDS JP 2001504324-A/19.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Gillespie,P.P., Kinoshita,C.M., Treco,D.A. and Williams,M.D.
TITLE Therapy for alpha-galactosidase A deficiency
JOURNAL Patent: JP 2001504324-A 19 03-APR-2001;
COMMENT TRANSKARYOTIC THERAPIES INC
PN JP 2001504324-A/19
PD 03-APR-2001
PF 12-SEP-1997 JP 1998514004
PR 13-SEP-1996 US 08/712614
PI RICHARD F SELDEN,MARIANNE BOROWSKI,FRANCES
P GILLESPIE,CAROL M
PI KINOSHITA,
PI DOUGLAS A TRECO,MELANIE D WILLIAMS
PC C12N9/40,C12N15/62,C12N15/56,A61K38/47,A61K48/00,C12N5/10 CC
Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..210
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x BD064230 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAGTACATAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 3
BD247363
LOCUS BD247363 210 bp DNA linear PAT 17-JUL-2003
DEFINITION Treatment for alpha-galactosidase A deficiency.
ACCESSION BD247363
VERSION BD247363.1 GI:33057133
KEYWORDS JP 2002538183-A/1.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 210)
AUTHORS Selden,R.F., Borowski,M., Kinoshita,C.M., Treco,D.A., Williams,M.D., Schuetz,T.J. and Daniel,P.F.
TITLE Treatment for alpha-galactosidase A deficiency
JOURNAL Patent: JP 2002538183-A 1 12-NOV-2002;

TRANSKARYOTIC THERAPIES INC
OS Artificial Sequence
PN JP 2002538183-A/1
PD 12-NOV-2002
PF 09-MAR-2000 JP 2000603353
PR 11-MAR-1999 US 09/266014
PI RICHARD F SELDEN,MARIANNE BOROWSKI,CAROL
M KINOSHITA,DOUGLAS A
PI TRECO,
PI MELANIE D WILLIAMS,THOMAS J SCHUETZ,PETER F DANIEL,PC
A61K38/43,A61P3/00,A61P3/08,C12N9/24/C12N15/09,A61K37/48, PC
C12N15/00
CC Description of Artificial Sequence: Human fibroblast library
CC probe: exon
7, including amplification primers.
FH Key Location/Qualifiers
FT source
1..210
/organism="Artificial Sequence".
FEATURES
source
1..210
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x BD247363 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAGTACATAATCCACAGGCACTGTTTGCCTTCAGCTA 208

RESULT 4
CS105028
LOCUS CS105028 210 bp DNA linear PAT 10-JUN-2005
DEFINITION Sequence 19 from Patent EP1538202.
ACCESSION CS105028
VERSION CS105028.1 GI:67512960
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unclassified.
REFERENCE 1
AUTHORS Selden,R.F., Borowski,M., Treco,D.A., Gillespie,P.P., Kinoshita,C.M. and Williams,M.D.
TITLE Production of human alpha-galactosidase A
JOURNAL Patent: EP 1538202-A 19 08-JUN-2005;
Transkaryotic Therapies, Inc. (US)
FEATURES
source
1..210
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x CS105028 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 155 ACTTCAAGTTAAGAGTACATAATCCACAGGCACTGTTTGCCTTCAGCTA 208

Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208
RESULT 5
LOCUS AR234437 210 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6458574.
ACCESSION AR234437
VERSION AR234437.1 GI:27277133
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Selden,R.F., Borowski,M., Kinoshita,C.M., Treco,D.A.,
Williams,M.D., Schuetz,T.J. and Daniel,P.F.
TITLE Treatment of a .alpha.-galactosidase deficiency
JOURNAL Patent: US 6458574-A 1 01-OCT-2002;
Transkaryotic Therapies, Inc.; Cambridge, MA
FEATURES
source
1. .210
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR234437 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208
RESULT 6
LOCUS AR322463 210 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 19 from patent US 6566099.
ACCESSION AR322463
VERSION AR322463.1 GI:33708228
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Selden,R.F., Borowski,M., Gillispie,F.P., Kinoshita,C.M.,
Treco,D.A. and Williams,M.D.
TITLE Nucleic acid encoding a chimeric polypeptide
JOURNAL Patent: US 6566099-A 19 20-MAY-2003;
Transkaryotic Therapies, Inc.; Cambridge, MA
FEATURES
source
1. .210
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR322463 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18

Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208
RESULT 7
LOCUS AR372139 210 bp DNA linear PAT 12-SEP-2003
DEFINITION Sequence 19 from patent US 6395884.
ACCESSION AR372139
VERSION AR372139.1 GI:34609448
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS Selden,R.F., Borowski,M., Gillispie,F.P., Kinoshita,C.M.,
Treco,D.A. and Williams,M.D.
TITLE Therapy for .alpha.-galactosidase deficiency
JOURNAL Patent: US 6395884-A 19 28-MAY-2002;
Transkaryotic Therapies, Inc.; DE
FEATURES
source
1. .210
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 3.69e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x AR372139 (1-210)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208
RESULT 8
LOCUS CQ684109 227 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 29035 from Patent WO2070737.
ACCESSION CQ684109
VERSION CQ684109.1 GI:42208348
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE Compositions and methods relating to osteoarthritis
JOURNAL Patent: WO 02070737-A 29035 12-SEP-2002;
Chondrogene Inc. (CA)
FEATURES
source
1. .227
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 4.01e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-29 (1-18) x CQ684109 (1-227)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18

```

Db      144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 197
|||||
RESULT 9
LOCUS      CQ685605                      270 bp      DNA      linear      PAT 03-FEB-2004
DEFINITION Sequence 30531 from Patent WO20070737.
ACCESSION  CQ685605
VERSION     CQ685605.1 GI:42213120
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1
AUTHORS    Liew,C.C., Marshall,W.E. and Zhang,H.
TITLE      Compositions and methods relating to osteoarthritis
JOURNAL    Patent: WO 02070737-A 30531 12-SEP-2002;
            Chondrogene Inc. (CA)
FEATURES    source
            location/Qualifiers
            1..270
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      4.82e-07      Length:      270
Score:          89.00        Matches:    18
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              6           Gaps:      0

US-10-602-220-29 (1-18) x CQ685605 (1-270)

Qy      1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db      174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 227
|||||

RESULT 10
LOCUS      I30314                      297 bp      DNA      linear      PAT 06-FEB-1997
DEFINITION Sequence 11 from patent US 5580757.
ACCESSION  I30314
VERSION     I30314.1 GI:1821105
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 297)
AUTHORS    Desnick,R.J., Bishop,D.F. and Ioannou,Y.A.
TITLE      Cloning and expression of biologically active .alpha.-galactosidase
            A as a fusion protein
JOURNAL    Patent: US 5580757-A 11 03-DEC-1996;
            Location/Qualifiers
FEATURES    source
            1..297
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      5.34e-07      Length:      297
Score:          89.00        Matches:    18
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              6           Gaps:      0

US-10-602-220-29 (1-18) x I30314 (1-297)

Qy      1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

```

```

Db      7 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 60
|||||
RESULT 11
LOCUS      AX030666                      297 bp      DNA      linear      PAT 20-SEP-2000
DEFINITION Sequence 8 from Patent EP1020528.
ACCESSION  AX030666
VERSION     AX030666.1 GI:10278187
KEYWORDS    .
SOURCE      unidentified
            unclassified sequences.
ORGANISM    unidentified
            unclassified sequences.
REFERENCE   1
AUTHORS    Ioannou,Y.A., Bishop,D.F. and Deanick,R.J.
TITLE      Method for producing secreted proteins
JOURNAL    Patent: EP 1020528-A 8 19-JUL-2000;
            SINAI SCHOOL MEDICINE (US)
FEATURES    Location/Qualifiers
            source
            1..297
            /organism="unidentified"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32644"
            <1..279
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAC09606.1"
            /db_xref="GI:10278188"
            /translation="EWTSRLRSHINPTGTVLLQLNTWQMSLKDLLPAGPAQHDEAAQQ
            NAFYQVLNMPNLNADQRNGFTQSLKDDPSOSANVLGEAQKLNDSQAPK"

ORIGIN
Alignment Scores:
Pred. No.:      5.34e-07      Length:      297
Score:          89.00        Matches:    18
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              6           Gaps:      0

US-10-602-220-29 (1-18) x AX030666 (1-297)

Qy      1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db      7 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 60
|||||

RESULT 12
LOCUS      ARI43996                      1158 bp     DNA      linear      PAT 08-AUG-2001
DEFINITION Sequence 53 from patent US 6210666.
ACCESSION  ARI43996
VERSION     ARI43996.1 GI:15105863
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1158)
AUTHORS    Miyamura,N.
TITLE      Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL    Patent: US 6210666-A 53 03-APR-2001;
            Location/Qualifiers
FEATURES    source
            1..1158
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Alignment Scores:
Pred. No.:      2.28e-06      Length:      1158
Score:          89.00        Matches:    18
Percent Similarity: 100.00%    Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    100.00%      Indels:    0
DB:              6           Gaps:      0

```

US-10-602-220-29 (1-18) x ARI43996 (1-1158)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 1158

RESULT 13

LOCUS ARI43995 1161 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 52 from patent US 6210666.
ACCESSION ARI43995
VERSION ARI43995.1 GI:15105862
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1161)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 52 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1161
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.28e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43995 (1-1161)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 1158

RESULT 14

LOCUS ARI43994 1164 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 51 from patent US 6210666.
ACCESSION ARI43994
VERSION ARI43994.1 GI:15105861
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1164)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 51 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1164
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.29e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43994 (1-1164)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 1158

RESULT 15

LOCUS ARI43993 1167 bp DNA linear PAT 08-AUG-2001
DEFINITION Sequence 50 from patent US 6210666.
ACCESSION ARI43993
VERSION ARI43993.1 GI:15105860
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Miyamura,N.
TITLE Truncated .alpha.-galactosidase A to treat fabry disease
JOURNAL Patent: US 6210666-A 50 03-APR-2001;
FEATURES Location/Qualifiers
source 1..1167
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 2.29e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x ARI43993 (1-1167)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||

Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTCTTCAGCTA 1158

Search completed: January 1, 2006, 03:35:28

Job time : 412.772 secs

This Page Blank (uspto)

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New).

Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: December 31, 2005, 23:48:43 ; Search time 48.2483 Seconds
(without alignments)
2486.399 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSLRLSHNPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn_1/USPTO.spool/US10602220/runat_30122005_140557_5411/app_query.fasta.1.782
-DB=N Geneseq -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220 @CGN 1.1.900 @runat_30122005_140557_5411 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-DEV_TIMEOUT=110 -WARN_TIMEOUT=30 -DELOP=6 -DELEXT=7
-FPGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	210	2	AAV31800 Nucleotid
2	89	100.0	210	3	AAA75432 Nucleotid
3	89	100.0	210	3	AAA70681 Human alp
4	89	100.0	210	6	AAD42074 Probe use

5	89	100.0	210	9	ACA62301	Ac62301 Probe for
6	89	100.0	210	10	AAD60098	Aad60098 210 bp pr
7	89	100.0	210	10	ADH00941	Adh00941 Probe fro
8	89	100.0	210	10	ADH54465	Adh54465 Probe to
9	89	100.0	210	12	ADG47825	Adg47825 Probe use
10	89	100.0	297	2	AAQ66242	Aaq66242 Protein A
11	89	100.0	1158	4	AAC85748	Aac85748 Mutant de
12	89	100.0	1161	4	AAC85747	Aac85747 Mutant de
13	89	100.0	1164	4	AAC85746	Aac85746 Mutant de
14	89	100.0	1167	4	AAC85745	Aac85745 Mutant de
15	89	100.0	1170	4	AAC85744	Aac85744 Mutant de
16	89	100.0	1173	4	AAC85743	Aac85743 Mutant de
17	89	100.0	1176	4	AAC85742	Aac85742 Mutant de
18	89	100.0	1179	4	AAC85741	Aac85741 Mutant de
19	89	100.0	1182	4	AAC85740	Aac85740 Mutant de
20	89	100.0	1188	4	AAC85739	Aac85739 Mutant de
21	89	100.0	1194	4	AAC85738	Aac85738 Alpha-gal
22	89	100.0	1197	3	AA70700	Aa70700 Human alp
23	89	100.0	1197	6	AAD42079	Aad42079 Human alp
24	89	100.0	1197	9	ACA62304	Ac62304 CDNA sequ
25	89	100.0	1197	10	AAD60103	Aad60103 Human alp
26	89	100.0	1197	10	ADH00921	Adh00921 Human alp
27	89	100.0	1197	10	ADH54443	Adh54443 Human alp
28	89	100.0	1197	12	ADG47805	Adg47805 Human alp
29	89	100.0	1214	11	ADI31675	Adi31675 Human con
30	89	100.0	1214	13	ADS83742	Ad83742 Human lym
31	89	100.0	1254	6	AAD45225	Aad45225 Human rga
32	89	100.0	1254	10	ADD84754	Add84754 Human alp
33	89	100.0	1254	12	ADM48684	Adm48684 Human wil
34	89	100.0	1254	13	ADU66919	Adu66919 Human alp
35	89	100.0	1254	14	AEA27448	Aea27448 Human alp
36	89	100.0	1266	6	AAD45223	Aad45223 Human rga
37	89	100.0	1266	10	ADD84750	Add84750 Human alp
38	89	100.0	1266	12	ADJ88276	Adj88276 Human wt
39	89	100.0	1266	12	ADM48680	Adm48680 Human wil
40	89	100.0	1266	13	ADU66915	Adu66915 Human alp
41	89	100.0	1266	14	AEA27444	Aea27444 Human alp
42	89	100.0	1272	6	AAD45226	Aad45226 Human rga
43	89	100.0	1272	10	ADD84756	Add84756 Human alp
44	89	100.0	1272	12	ADJ88282	Adj88282 Human wt
45	89	100.0	1272	12	ADM48686	Adm48686 Human wil

ALIGNMENTS

RESULT 1
AAV31800
ID AAV31800 standard; DNA; 210 BP.
XX
AC AAV31800;
XX
DT 21-AUG-1998 (first entry)
XX
DE Nucleotide sequence of the human alpha-gal A gene probe.
XX
KW Probe; hybridisation; human alpha-galactosidase A gene; alpha-gal A gene;
KW glycosylated human alpha-gal A; Fabry disease; ss;
KW alpha-gal A deficiency; Fabry; X-linked; lysosomal storage disease.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO9811206-A2.
XX
PD 19-MAR-1998.
XX
PF 12-SEP-1997; 97WO-US016603.
XX
PR 13-SEP-1996; 96US-00712614.
XX
PA (TRAN-) TRANSKARYOTIC THERAPIES INC.
PI Seiden RF, Borowski M, Gillespie FP, Kinoshita CM, Treco DA;

PI Williams MD;
 XX WPI; 1998-207375/18.
 XX
 XX Treatment of alpha-galactosidase A deficiency in, e.g. Fabry disease -
 PT comprises use of genetically modified human cells that express alpha-
 PT galactosidase.
 XX
 XX Disclosure; Fig 1; 78pp; English.
 XX
 XX This is the nucleotide sequence of the probe used in the to isolate the
 CC human alpha-galactosidase A gene (alpha-gal A), used in the method of the
 CC invention, which involves the production of genetically modified human
 CC cells that express alpha-galactosidase. The cultured human cells can be
 CC used for the production of glycosylated human alpha-gal A and can be used
 CC for the treatment of patients with an alpha-gal A deficiency, e.g. Fabry
 CC disease (an X-linked inherited lysosomal storage disease)
 XX
 XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0
 US-10-602-220-29 (1-18) x AAV31800 (1-210)
 QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTTCAGCTA 208
 RESULT 2
 AAA75432
 ID AAA75432 standard; DNA; 210 BP.
 XX
 AC AAA75432;
 XX
 XX 15-JAN-2001 (first entry)
 DT
 XX Nucleotide sequence of a probe from exon 7 of human alpha-Gal A gene.
 DE
 XX Human; alpha-galactosidase A; alpha-Gal A; lysosomal glycohydrolase;
 XX Fabry's disease; cardiovascular abnormality; probe;
 KW left ventricular hypertrophy; LVH; ss.
 XX
 XX Synthetic.
 OS
 OS Homo sapiens.
 XX
 XX WO200053730-A2.
 PN
 XX 14-SEP-2000.
 PD
 XX
 XX 09-MAR-2000; 2000WO-US006118.
 PF
 XX
 XX 11-MAR-1999; 99US-00266014.
 PR
 XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
 PA
 XX Selden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;
 PI Schuetz TJ, Daniel PF;
 PI
 XX WPI; 2000-594316/56.
 DR
 XX Pure human alpha-Gal A preparation for treating alpha-galactosidase A
 PT deficiency associated disorders e.g. Fabry's disease, cardiovascular
 PT abnormality, has various alpha-Gal A glycoforms and is free of lectins.
 XX
 XX Example 1; Fig 1; 92pp; English.
 PS
 XX The present sequence represents a probe derived from exon 7 of the human

CC alpha-galactosidase A (alpha-Gal A) gene. Alpha-Gal A is lysosomal
 CC glycohydrolase which cleaves the terminal alpha-galactosyl moieties of
 CC various glycoconjugates. The protein is useful for preparing a medicament
 CC for treating alpha-Gal A deficiency associated disorders, e.g. typical
 CC variant of Fabry's disease, cardiovascular abnormality, preferably left
 CC ventricular hypertrophy (LVH). Alpha-Gal A produced by the invention has
 CC an extended circulating half-life and increased uptake in specific
 CC tissues other than live, compared with alpha-Gal A produced by other
 CC methods
 XX
 XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0
 US-10-602-220-29 (1-18) x AAA75432 (1-210)
 QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 DB 155 ACTTCAAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTTCAGCTA 208
 RESULT 3
 AAA70681
 ID AAA70681 standard; DNA; 210 BP.
 XX
 AC AAA70681;
 XX
 XX 31-JAN-2001 (first entry)
 DT
 XX Human alpha-Gal A gene exon 7 probe.
 DE
 XX Nephrotropic; gene therapy; alpha-galactosidase A deficiency; ss;
 KW Fabry disease; secretion; human growth hormone; hGH; signal peptide;
 KW enzyme replacement therapy; probe hybridisation.
 XX
 XX Homo sapiens.
 OS
 XX US6083725-A.
 PN
 XX 04-JUL-2000.
 PD
 XX 12-SEP-1997; 97US-00928881.
 PF
 XX 13-SEP-1996; 96US-0026041P.
 PR
 XX (TRAN-) TRANSKARYOTIC THERAPIES INC.
 PA
 XX Kinoshita CM, Treco DA, Williams MD, Selden RF, Borowski M;
 PI Gillispie FP;
 PI
 XX WPI; 2000-464341/40.
 DR
 XX New DNA comprising a human growth hormone signal peptide and human alpha-
 PT galactosidase A polypeptide and cells expressing the DNA, useful for
 PT treating Fabry disease and in vitro protein production for enzyme
 PT replacement therapy.
 XX
 XX Example 1; Fig 1; 32pp; English.
 PS
 XX The invention relates to the treatment of a patient having an alpha-
 CC galactosidase A (aga) deficiency, e.g. Fabry disease, by using cells
 CC overexpressing and secreting aga or purified aga. This sequence
 CC represents a 270 bp fragment of the human aga exon 7 sequence for use as
 CC a probe to isolate the complete aga gene sequence (AAA70685). The probe
 CC was amplified by the primers AAA70679-A70680. For expression and
 CC secretion of the aga, the protein is fused to a signal peptide especially
 CC the human growth hormone (hGH) signal peptide. Primary, secondary, or
 CC immortalised human cells genetically modified to overexpress human aga

CC are also useful for in vitro protein production, to produce protein which
 CC may be purified for enzyme replacement therapy. The method allows the
 CC production of large quantities of appropriately glycosylated and
 CC therapeutically useful human alpha

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x AAA70681 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
 DB 155 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 208

RESULT 4

AAD42074
 ID AAD42074 standard; DNA; 210 BP.

XX
 AC AAD42074;

XX
 DT 04-NOV-2002 (first entry)

XX
 DE Probe used to isolate human alpha-gal A cDNA.

XX
 KW Human; alpha-galactosidase A; alpha-gal A deficiency; Fabry disease;
 KW enzyme; probe; ss.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT primer_bind complement(1..21)
 FT /*tag= a
 FT /bound_moiety= "Primer 1"
 FT primer_bind 190..210
 FT /*tag= b
 FT /bound_moiety= "Primer 2"

XX
 PN US6395884-B1.

XX
 PD 28-MAY-2002.

XX
 PF 06-APR-2000; 2000US-00543921.

XX
 PR 13-SEP-1996; 96US-0026041P.

XX
 PR 12-SEP-1997; 97US-00928881.

XX
 PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX
 PI Seiden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;
 PI Williams MD;

XX
 DR WPI; 2002-526374/56.

XX
 PT Purifying human alpha-galactosidase A from a cell for use in treating
 PT Fabry disease, by passing sample containing the enzyme over hydrophobic
 PT interaction resin, heparin resin, hydroxyapatite and anion exchange
 PT resin.

XX
 PS Example 1; Col 31-32; 31pp; English.

XX
 CC The invention relates to a method of purifying human alpha-galactosidase
 CC A (alpha-gal A) from a human cell. The method involves a chromatography
 CC step which comprises passing human alpha-gal A containing sample over a
 CC hydrophobic interaction resin. The method is useful for purifying alpha-
 CC gal A from a human cell, for use in treating an individual suspected of
 CC having alpha-gal A deficiency such as Fabry disease. The present sequence

CC is a probe used to isolate human alpha-gal A cDNA

XX Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x AAD42074 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
 DB 155 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 208

RESULT 5

ACA62301

ID ACA62301 standard; DNA; 210 BP.

XX
 AC ACA62301;

XX
 DT 13-AUG-2003 (first entry)

XX
 DE Probe for cDNA encoding human alpha-galactosidase A glycoprotein.

XX
 KW Human; alpha-galactosidase A glycoprotein; alpha-gal A; glycan;
 KW complex-type glycan; alpha-galactosidase A deficiency; Fabry's disease;
 KW cardiovascular abnormality; ventricular enlargement;
 KW left ventricular hypertrophy; mitral valve insufficiency;
 KW renal abnormality; sialylation; nephrotropic; probe; ss.

XX
 OS Homo sapiens.

XX
 PN US6458574-B1.

XX
 PD 01-OCT-2002.

XX
 PF 11-MAR-1999; 99US-00266014.

XX
 PR 13-SEP-1996; 96US-0026041P.

XX
 PR 12-SEP-1997; 97US-00928881.

XX
 PR 12-SEP-1997; 97WO-US016603.

XX
 PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX
 PI Seiden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;

XX
 PI Schuetz TJ, Daniel PF;

XX
 DR WPI; 2003-465541/44.

XX
 PT Human alpha-galactosidase A glycoprotein preparation containing complex-
 PT type glycans, useful for treating alpha-galactosidase A deficiency,
 PT especially Fabry's disease.

XX
 PS Example 1; Fig 1; 39pp; English.

XX
 CC The present invention relates to the isolation of highly purified human
 CC alpha-galactosidase A (alpha-gal A) glycoprotein, and preparations
 CC containing alpha-gal A, where greater than 50% of the total glycans of
 CC the preparation are complex-type glycans, produced by providing a human
 CC cell genetically modified to express alpha-gal A polypeptide, and
 CC purifying the polypeptide from the human cell or its culture medium. The
 CC alpha-gal A preparations are useful for the treatment of alpha-
 CC galactosidase A deficiency, especially Fabry's disease e.g. specific
 CC populations of Fabry patients with predominantly cardiovascular
 CC abnormalities such as ventricular enlargement (e.g. left ventricular
 CC hypertrophy and/or mitral valve insufficiency), or Fabry patients with
 CC predominantly renal involvement. The alpha-gal A preparations have
 CC extended circulating half-life in mammalian hosts, as improved
 CC sialylation of alpha-gal A preparations enhances the circulatory half-

CC life of exogenous alpha-Gal A. The present sequence represents a probe
 CC used to isolate cDNA encoding human alpha-Gal A in the examples of the
 CC present invention

XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0

US-10-602-220-29 (1-18) x ACA62301 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208

RESULT 6

AD60098
 ID AAD60098 standard; DNA; 210 BP.

XX AC AAD60098;

XX DT 18-DEC-2003 (first entry)

XX DE 210 bp probe used to isolate human alpha-gal A gene.

XX KW Human growth hormone; glucocerebrosidase; gene therapy; cardiovascular;
 XX KW Gaucher's disease; Fabry disease; alpha-galactosidase A deficiency; hGH;
 XX KW CNS; nephrotropic; human; probe; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT primer_bind complement(1..22)
 XX FT /*tag= a
 XX FT primer_bind /bound_moiety= "primer 1"
 XX FT 190..210
 XX FT /*tag= b
 XX FT /bound_moiety= "primer 2"

XX US5566099-B1.

XX PD 20-MAY-2003.

XX PF 27-JAN-2000; 2000US-00491759.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PA Seiden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;
 XX PI Williams MD;

XX WPI; 2003-719643/68.

XX Novel polypeptide comprising signal peptide of human growth hormone
 XX linked to glucocerebrosidase useful for treating disorders associated
 XX with enzyme deficiency in sphingolipid degradation such as gaucher
 XX disease.

XX Example 1; Fig 1; 30pp; English.

XX The present invention relates to a novel polypeptide comprising signal
 XX peptide of human growth hormone (hGH) linked to glucocerebrosidase. The
 XX invention is useful for treating disorders associated with alpha-
 XX galactosidase A deficiency and sphingolipid degradation such as Gaucher's
 XX disease and Fabry disease. The invention is also useful in gene therapy.
 XX The present sequence is a 210 bp probe used to isolate human alpha-gal A

CC gene

XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x AAD60098 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208

RESULT 7

ADH00941
 ID ADH00941 standard; DNA; 210 BP.

XX AC ADH00941;

XX DT 11-MAR-2004 (first entry)

XX DE Probe from human alpha-galactosidase A (alpha-Gal A) gene.

XX KW Human; ss; alpha-galactosidase A; alpha-Gal A; enzyme therapy; therapy;
 XX KW Fabry disease; cardiovascular abnormality; left ventricular hypertrophy;
 XX KW cardiact; probe.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 XX FT primer_bind 1..22
 XX FT /*tag= a
 XX FT primer_bind 190..210
 XX FT /*tag= b

XX US2003077806-A1.

XX PD 24-APR-2003.

XX PF 07-JUN-2002; 2002US-00165060.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 12-SEP-1997; 97WO-US016603.

XX PR 11-MAR-1999; 99US-00266014.

XX (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PA Seiden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;
 XX PI Schuetz TJ, Daniel PF;

XX WPI; 2003-567381/53.

XX Composition useful for treating Fabry disease or atypical variant of
 XX Fabry disease comprising a human alpha-galactosidase A preparation,
 XX purified to at least 98% homogeneity.

XX Example 1; SEQ ID NO 1; 42pp; English.

XX The invention relates to a composition comprising a human alpha-
 XX galactosidase A (alpha-Gal A) preparation, purified to at least 98%
 XX homogeneity, as measured by sodium dodecyl sulphate polyacrylamide gel
 XX electrophoresis (SDS-PAGE) or reverse phase HPLC, where the human alpha-
 XX Gal A preparation has a specific activity of at least 2.0x10⁶
 XX units/mg protein. The invention is useful in enzyme therapy. The
 XX invention is useful for treating a patient with Fabry disease, or for
 XX treating a patient with an atypical variant of Fabry disease (e.g.
 XX cardiovascular abnormality such as left ventricular hypertrophy), where

CC the alpha-Gal A preparation is administered to the patient in a dose
 CC between 0.05-5 mg of alpha-Gal A preparation per kg body weight weekly or
 CC biweekly. The present sequence is a probe from human alpha-Gal A gene
 CC used to isolate the cDNA.

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x ADH00941 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 DB 155 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTTCAGCTA 208

RESULT 8

ADH54465
 ID ADH54465 standard; cDNA; 210 BP.

XX AC ADH54465;

XX DT 25-MAR-2004 (first entry)

XX DE Probe to isolate human alpha-galactosidase A (alpha-gal A) cDNA.

XX KW Alpha-galactosidase A; nephrotropic; nephrotropic; gene therapy;

XX KW Fabry disease; human; ss; probe.

XX OS Homo sapiens.

XX PN US2003152560-A1.

XX PD 14-AUG-2003.

XX PF 12-DEC-2002; 2002US-00318905.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 27-JAN-2000; 2000US-00491759.

XX PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PI Selden RF, Borowski M, Gillispie FP, Kinoshita CM, Treco DA;

XX PI Williams MD;

XX DR WPI; 2003-897682/82.

XX PT Therapeutic method in alpha-galactosidase A deficiency using genetically

XX PT modified cells overexpressing and secreting the human alpha-galactosidase

XX PT A, useful particularly for treating Fabry disease.

XX PS Example 1; SEQ ID NO 19; 33pp; English.

XX CC The invention relates to a method of treatment which comprises

XX CC identifying a patient suspected of having an alpha-galactosidase A

XX CC deficiency, and introducing into the patient a human cell genetically

XX CC modified to overexpress and secrete human alpha-gal A. The invention is

XX CC useful in gene therapy. The methods and compositions of the present

XX CC invention are useful for treating disorders with alpha-galactosidase

XX CC deficiency, such as Fabry disease. The present sequence is a probe used

XX CC to isolate alpha-galactosidase A cDNA.

SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.41e-07 Length: 210
 Score: 89.00 Matches: 18

Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-10-602-220-29 (1-18) x ADH54465 (1-210)

QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 DB 155 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTTCAGCTA 208

RESULT 9

ADG47825

ID ADG47825 standard; DNA; 210 BP.

XX AC ADG47825;

XX DT 11-MAR-2004 (first entry)

XX DE Probe used to isolate human alpha-Gal A cDNA.

XX KW Alpha-galactosidase A; alpha-Gal A; Fabry disease;
 KW cardiovascular abnormality; ventricular enlargement;
 KW left ventricular hypertrophy; LVH; mitral valve insufficiency;
 KW renal involvement; enzyme therapy; human; probe; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT primer_bind complement(1..22)

XX FT primer_bind /*tag= a

XX FT primer_bind 190..210

XX FT primer_bind /*tag= b

XX PN US2003113894-A1.

XX PD 19-JUN-2003.

XX PF 10-JUN-2002; 2002US-00165968.

XX PR 13-SEP-1996; 96US-0026041P.

XX PR 12-SEP-1997; 97US-00928881.

XX PR 12-SEP-1997; 97WO-US016603.

XX PR 11-MAR-1999; 99US-00266014.

XX PA (TRAN-) TRANSKARYOTIC THERAPIES INC.

XX PI Selden RF, Borowski M, Kinoshita CM, Treco DA, Williams MD;

XX PI Schuetz TJ, Daniel PF;

XX DR WPI; 2004-009047/01.

XX CC Composition useful for treating Fabry disease or atypical variant of
 XX CC Fabry disease comprising a human alpha-galactosidase A preparation,
 XX CC purified to at least 98 percent homogeneity.

XX PS Example 1; SEQ ID NO 1; 43pp; English.

XX CC The present invention provides highly purified alpha-galactosidase A

XX CC (alpha-Gal A) and various methods for purifying it; alpha-Gal A

XX CC preparations with altered charge and methods for making those

XX CC preparations; alpha-Gal A preparations that have an extended circulating

XX CC half-life in a mammalian host and methods for making the same; and

XX CC methods and dosages for administering an alpha-Gal preparation to a

XX CC subject. The alpha-Gal A preparation of the invention is useful for

XX CC treating a patient with Fabry disease, or atypical variants of Fabry

XX CC disease e.g. cardiovascular abnormality such as ventricular enlargement

XX CC e.g. left ventricular hypertrophy (LVH) and/or mitral valve insufficiency

XX CC or Fabry patients with predominantly renal involvement. The present

XX CC sequence is a probe used to isolate human alpha-Gal A cDNA.

XX SQ Sequence 210 BP; 52 A; 51 C; 52 G; 55 T; 0 U; 0 Other;

```
Alignment Scores:
Pred. No.:      8.41e-07      Length:      210
Score:          89.00        Matches:      18
Percent Similarity: 100.00%  Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     100.00%    Indels:      0
DB:              12         Gaps:      0

US-10-602-220-29 (1-18) x ADG47825 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 208

RESULT 10
AAQ66242
ID AAQ66242 standard; DNA; 297 BP.
XX
AC AAQ66242;
XX
DT 25-MAR-2003 (revised)
DT 15-NOV-1994 (first entry)
XX
DE Protein A domain E.
XX
KW Alpha-galactosidase A; Fabry disease; blood group O; blood group B;
KW COS-1; enzyme replacement therapy; glycoconjugate; protein A;
KW IgG binding domain; fusion protein; alpha-Gal A; ds.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT CDS 1..279
FT /tag= a
XX
XX WO9412628-A1.
XX
XX 09-JUN-1994.
XX
XX 30-NOV-1993; 93WO-US011539.
XX
XX 30-NOV-1992; 92US-00983451.
XX
XX (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
XX Deanick RJ, Bishop DF, Ioannou YA;
XX
XX WPI; 1994-200257/24.
XX
XX P-PSDB; AAR53771.
XX
XX Prodn of human alpha-galactosidase A - by culturing cells contg the
XX coding sequence and the beta-galactosidase alpha-2,6-sialyl transferase
XX gene and regulation sequences.
XX
XX Disclosure; Page 107-108; 156pp; English.
XX
XX The cDNA sequence (AAQ66241) of human alpha-galactosidase A (alpha-Gal A)
XX was determined. A fusion protein of alpha-Gal A cDNA and protein A IGG
XX binding domain E (AAQ66242, AAR53771) was expressed in COS-1 cells. The
XX alpha-Gal A cDNA was amplified by PCR using primers (AAQ66243-44)
XX designed to delete the termination codon, to add a 3' collagenase
XX cleavage consensus sequence and to include a 5' EcoRI recognition site.
XX The protein A domain E was amplified with the collagenase consensus
XX sequence in the 5' oligonucleotide using the primers given in AAQ66245-
XX 46. The enzyme can be used for Fabry disease enzyme replacement therapy,
XX to convert blood group B to O, or to hydrolyze alpha-D-galactosyl
XX residues from glycoconjugates. (Updated on 25-MAR-2003 to correct EN
XX field.)
XX
XX Sequence 297 BP; 103 A; 64 C; 54 G; 76 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      1.29e-06      Length:      297

US-10-602-220-29 (1-18) x AAQ66242 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 7 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 60

RESULT 11
AAC85748
ID AAC85748 standard; cDNA; 1158 BP.
XX
AC AAC85748;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 12 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
XX Homo sapiens.
XX
XX US6210666-B1.
XX
XX 03-APR-2001.
XX
XX 21-OCT-1998; 98US-00176666.
XX
XX 21-OCT-1997; 97US-0062560P.
XX
XX (ORPH-) ORPHAN MEDICAL INC.
XX
XX Miyamura N;
XX
XX WPI; 2001-289627/30.
XX
XX P-PSDB; AAB47203.
XX
XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
XX in humans or for treating a condition associated with a deficiency of
XX alpha-galactosidase A.
XX
XX Disclosure; Col 71-74; 51pp; English.
XX
XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
XX which has a C-terminal truncation of 12 amino acids. Alpha-Gal A or its
XX variant, which has a carboxy-terminal deletion of 2-11 amino acids and
XX which exhibits alpha-galactosidase A enzyme activity, is useful for
XX treating Fabry disease or a condition associated with a deficiency of
XX alpha-galactosidase A. Fabry disease has been linked to an insertion of
XX an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
XX sequence, which results in a tyrosine-to-stop codon substitution at amino
XX acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX
XX Sequence 1158 BP; 299 A; 259 C; 301 G; 299 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      6.85e-06      Length:      1158
Score:          89.00        Matches:      18
Percent Similarity: 100.00%  Conservaive: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:     100.00%    Indels:      0
DB:              4         Gaps:      0

US-10-602-220-29 (1-18) x AAC85748 (1-1158)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 1158
```

RESULT 12
AAC85747
ID AAC85747 standard; cDNA; 1161 BP.
XX
AC AAC85747;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 11 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
PN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX
PR 21-OCT-1997; 97US-0062560P.
XX
PA (ORPH-) ORPHAN MEDICAL INC.
XX
PI Miyamura N;
XX
PI WPI; 2001-289627/30.
XX
DR P-PSDB; AAB47202.
XX
PT New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of
PT alpha-galactosidase A.
XX
PS Disclosure; Col 71-72; 51pp; English.
XX
CC This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 11 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX
SQ Sequence 1161 BP; 301 A; 259 C; 302 G; 299 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 6.87e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-602-220-29 (1-18) x AAC85747 (1-1161)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 1158
RESULT 13
AAC85746
ID AAC85746 standard; cDNA; 1164 BP.
XX
AC AAC85746;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 10 alpha-galactosidase A cDNA.
XX

KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
PN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX
PR 21-OCT-1997; 97US-0062560P.
XX
PA (ORPH-) ORPHAN MEDICAL INC.
XX
PI Miyamura N;
XX
PI WPI; 2001-289627/30.
XX
DR P-PSDB; AAB47201.
XX
PT New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of
PT alpha-galactosidase A.
XX
PS Disclosure; Col 69-72; 51pp; English.
XX
CC This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 10 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX
SQ Sequence 1164 BP; 303 A; 259 C; 302 G; 300 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 6.89e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
US-10-602-220-29 (1-18) x AAC85746 (1-1164)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTCAGCTA 1158
RESULT 14
AAC85745
ID AAC85745 standard; cDNA; 1167 BP.
XX
AC AAC85745;
XX
DT 18-JUL-2001 (first entry)
XX
DE Mutant delta 9 alpha-galactosidase A cDNA.
XX
KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
KW Fabry disease; C-terminal truncation; ss.
XX
OS Homo sapiens.
XX
PN US6210666-B1.
XX
PD 03-APR-2001.
XX
PF 21-OCT-1998; 98US-00176666.
XX

```

PR 21-OCT-1997; 97US-0062560P.
XX (ORPH-) ORPHAN MEDICAL INC.
XX Miyamura N;
XX WPI; 2001-289627/30.
XX P-PSDB; AAB47200.
XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of
PT alpha-galactosidase A.
XX Disclosure; Col 67-70; 51pp; English.
XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 9 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX SQ Sequence 1167 BP; 305 A; 260 C; 302 G; 300 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.91e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85745 (1-1167)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAGAGAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 15
AAC85744
ID AAC85744 standard; cDNA; 1170 BP.
XX AC AAC85744;
XX 18-JUL-2001 (first entry)
XX DE Mutant delta 8 alpha-galactosidase A cDNA.
XX KW Wild type; human; alpha-galactosidase A; alpha-Gal-A; variant; deletion;
XX KW Fabry disease; C-terminal truncation; ss.
XX OS Homo sapiens.
XX EN US6210666-B1.
XX PD 03-APR-2001.
XX PF 21-OCT-1998; 98US-00176666.
XX PR 21-OCT-1997; 97US-0062560P.
XX (ORPH-) ORPHAN MEDICAL INC.
XX PI Miyamura N;
XX WPI; 2001-289627/30.
XX P-ESDB; AAB47199.
XX New alpha-galactosidase A polypeptide, useful for treating Fabry disease
PT in humans or for treating a condition associated with a deficiency of

```

```

PT alpha-galactosidase A.
XX Disclosure; Col 67-68; 51pp; English.
XX This sequence encodes a variant human alpha-galactosidase A (alpha-Gal-A)
CC which has a C-terminal truncation of 8 amino acids. Alpha-Gal A or its
CC variant, which has a carboxy-terminal deletion of 2-11 amino acids and
CC which exhibits alpha-galactosidase A enzyme activity, is useful for
CC treating Fabry disease or a condition associated with a deficiency of
CC alpha-galactosidase A. Fabry disease has been linked to an insertion of
CC an adenine nucleotide at nt 1095 in exon 7 of the alpha-Gal A coding
CC sequence, which results in a tyrosine-to-stop codon substitution at amino
CC acid residue 365 (Y365X), causing a C-terminal truncation by 65 residues
XX SQ Sequence 1170 BP; 306 A; 260 C; 303 G; 301 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 6.94e-06 Length: 1170
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-29 (1-18) x AAC85744 (1-1170)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAGAGAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

Search completed: January 1, 2006, 00:49:55
Job time : 51.2483 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:17 ; Search time 326.524 Seconds
(without alignments)
2579.189 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSLRLSHINPTGTVLLQL 18

Scoring table: BLOSUM62 Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO spo01/US1060220/runat_30122005_140558_5433/app_query.fasta_1.782
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1060220 @CNC 1.1 6731 @runat_30122005_140558_5433 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_hic:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gse1:*
10: gb_gse2:*
11: gb_gse3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	89	100.0	177	1	AI621155
C 2	89	100.0	311	6	CD365384
C 3	89	100.0	319	2	BP061865
C 4	89	100.0	346	1	AA251784
C 5	89	100.0	358	1	AW948701
C 6	89	100.0	371	3	BI492391
C 7	89	100.0	372	1	AW021436

C 8	89	100.0	387	1	AI758588
C 9	89	100.0	393	1	AW004664
C 10	89	100.0	397	1	AA287599
C 11	89	100.0	398	1	AW449396
C 12	89	100.0	401	1	AI948576
C 13	89	100.0	403	1	AI949177
C 14	89	100.0	407	1	AA548321
C 15	89	100.0	427	1	AU149911
C 16	89	100.0	443	1	AW470828
C 17	89	100.0	445	1	AW135099
C 18	89	100.0	446	6	CD742642
C 19	89	100.0	460	6	CA777417
C 20	89	100.0	461	2	BF001466
C 21	89	100.0	462	2	BF111590
C 22	89	100.0	462	5	BU740166
C 23	89	100.0	481	1	AI799133
C 24	89	100.0	496	1	AI708415
C 25	89	100.0	499	1	AA708506
C 26	89	100.0	509	1	AA622854
C 27	89	100.0	512	1	AI859628
C 28	89	100.0	521	1	AA664321
C 29	89	100.0	524	2	BG232101
C 30	89	100.0	530	6	CB529984
C 31	89	100.0	542	1	AI469893
C 32	89	100.0	544	1	AW731746
C 33	89	100.0	547	2	BF590466
C 34	89	100.0	547	2	BE855457
C 35	89	100.0	547	2	BE855462
C 36	89	100.0	551	3	BP299850
C 37	89	100.0	554	1	AA463854
C 38	89	100.0	554	5	BU615633
C 39	89	100.0	568	3	BP299872
C 40	89	100.0	575	1	AI800468
C 41	89	100.0	577	1	AW517319
C 42	89	100.0	593	1	AW264095
C 43	89	100.0	600	6	CD367153
C 44	89	100.0	601	5	BU607709
C 45	89	100.0	602	1	AI859634

ALIGNMENTS

RESULT 1
AI621155/c

LOCUS
tu88b03.x1

DEFINITION
NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2258093 3', similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION
AI621155

VERSION
AI621155.1

KEYWORDS
GI:4630281

SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens

REFERENCE
1 (bases 1 to 177)

AUTHORS
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL
Unpublished (1997)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.
Location/Qualifiers

177 bp mRNA linear EST 21-APR-1999
tu88b03.x1 NCI CGAP Gas4 Homo sapiens cDNA clone IMAGE:2258093 3', similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);, mRNA sequence.

AI621155
AI621155.1 GI:4630281
EST.
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 177)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers

```

source
1. 177
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2258093"
/tissue type="poorly differentiated adenocarcinoma with
  signet ring cell features"
/lab host="DH10B"
/clone_lib="NCI CGAP Gas4"
/notes="Organ: stomach; Vector: pCMV-SPORT6; Site 1: SalI;
  Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
  Average insert size 1.69 kb. Life Technologies catalog #:
  11549-011"

ORIGIN
Alignment Scores:
Pred. No.: 2.8e-05 Length: 177
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x A1621155 (1-177)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 96 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 43

RESULT 2
LOCUS CD365384 311 bp mRNA linear EST 05-AUG-2004
DEFINITION UI-H-FT2-bj1-d-03-0-UI.s1 NCI CGAP FT2 Homo sapiens cDNA clone
similar to SW:AGAL_HUMAN P06280 ALPHA-GALACTOSIDASE A PRECURSOR ;,
mRNA sequence.
ACCESSION CD365384
VERSION CD365384.1 GI:31149474
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TUMOR Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. M. Bento Soares, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/cgap.html
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
Location/Qualifiers
1..311
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FT2-bj1-d-03-0-UI"
/tissue type="Alveolar Macrophage"
/dev stage="Adult"
/lab host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP FT2"
/notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP FT2 is a subtracted cDNA library constructed from
a pool of 81 RNA samples from Alveolar Macrophages
challenged with different treatments. The mRNA samples
were a mixture of these conditions (times refer to
incubations following isolation by bronchoalveolar lavage)
(some normal donor macrophages were cultured in some of
the conditions, other donor macrophages in different
conditions). The mRNA samples were pooled for library
construction. Control 0 hours; control 3 hours; control 24
hours; LPS 100 ng/ml, 3 hours; LPS 100 ng/ml, 24 hours;
PMA 10 ng/ml, 3 hours; PMA 10 ng/ml, 24 hours; Klebsiella
moi 10, 3 hours; Klebsiella moi 10, 24 hours; Staph
aureus moi 10, 3 hours; Staph aureus moi 10, 24 hours;
Adenoviral vector (Ad5 CMV eGFP), moi 500, 3 hours; wt
Adenoviral vector (Ad5 CMV eGFP), moi 500, 24 hours; wt
adenovirus moi 500, 3 hours; wt adenovirus moi 500, 24
hours; Ad vector + LPS 3 hours; Ad vector + LPS 24 hours;
wt adenovirus + LPS 3 hours; wt adenovirus + LPS 24 hours.
The library was subtracted according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. The tissue
was provided by Dr. Gary W. Hunninghake of the University
of Iowa.
TAG TISSUE=Human Lung Alveolar Macrophage
TAG LIB=UI-H-FT2
TAG_SEQ=GGCCATGCCG"

ORIGIN
Alignment Scores:
Pred. No.: 5.31e-05 Length: 311
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x CD365384 (1-311)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 109 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 56

RESULT 3
LOCUS BF061865 319 bp mRNA linear EST 16-OCT-2000
DEFINITION 7K68a08.x1 NCI CGAP GC6 Homo sapiens cDNA clone IMAGE:3480471.3
similar to SW:AGAL_HUMAN P06280 ALPHA-GALACTOSIDASE A PRECURSOR ;,
mRNA sequence.
ACCESSION BF061865
VERSION BF061865.1 GI:10820775
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 319)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TUMOR National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TUMOR Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco.
Location/Qualifiers
1..319
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source

```

/clone="IMAGE:3480471"
 /tissue type="pooled germ cell tumors"
 /lab host="DH10B"
 /clone lib="NCI CGAP GC6"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA was prepared from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaudo."

ORIGIN

Alignment Scores:
 Pred. No.: 5,46e-05 Length: 319
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x BF061865 (1-319)

Qy 1 ThrSerArgLeuArgSerHisIleAenProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 102 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGCTTCAGCTA 49

RESULT 4

AA251784/c
 LOCUS AA251784 346 bp mRNA linear EST 13-AUG-1997
 DEFINITION z811d08.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:684879 3' similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN); mRNA sequence.

ACCESSION

AA251784

VERSION

AA251784.1 GI:1886746

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 346)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

This clone is available royalty-free through LILNL; contact the IMAGE Consortium (info@image.lilnl.gov) for further information.

Seq primer: 41ml3 fwd. ET from Amersham

High quality sequence stop: 307.

Location/Qualifiers

1..346

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:684879"

/tissue type="germinal center B cell"

/lab host="DH10B"

/clone lib="NCI CGAP GCBI"

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer

[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-3'

). Double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

ORIGIN

Alignment Scores:
 Pred. No.: 5,99e-05 Length: 346
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AA251784 (1-346)

Qy 1 ThrSerArgLeuArgSerHisIleAenProThrGlyThrValLeuLeuGlnLeu 18
 |||||
 Db 100 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTGCTTCAGCTA 47

RESULT 5

AW948701

LOCUS

AW948701

DEFINITION

OVO-FT0002-050500-231-h02 FT0002 Homo sapiens cDNA, mRNA sequence.

ACCESSION

AW948701

VERSION

AW948701.1 GI:8126475

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 358)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

10737800

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL

(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=st2=QV0-FT0002-050>)

500-231-h02st3=2000-05-05st4=1)

Seq primer: puc 18 forward

High quality sequence start: 10

High quality sequence stop: 358.

Location/Qualifiers

1..358

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev stage="Adult"

/clone lib="FT0002"

/note="Organ: prostate tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaudo."

```

Alignment Scores:
Pred. No.: 6.23e-05 Length: 358
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW948701 (1-358)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 279 ACTTCAGGTTAAGAGTACATATATCCACAGGCACTGTTTGGCTTCAGCTA 332

RESULT 6
LOCUS BI492391/c 371 bp mRNA linear EST 28-AUG-2001
DEFINITION IMAGE:2484240 3', mRNA sequence.
ACCESSION BI492391
VERSION BI492391.1 GI:15331735
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 371)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
JOURNAL Genomics 23, 42-50 (1994)
PUBMED 7829101
COMMENT Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Plate: LLAM6181 row: L column: 1
Seq primer: P7 primer.
Location/Qualifiers
1. .371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2484240"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: ECoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"

FEATURES
source
1. .371
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2484240"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: ECoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"

ORIGIN
Alignment Scores:
Pred. No.: 6.49e-05 Length: 371
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

US-10-602-220-29 (1-18) x AW948701 (1-358)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 279 ACTTCAGGTTAAGAGTACATATATCCACAGGCACTGTTTGGCTTCAGCTA 332

RESULT 7
LOCUS AW021436 372 bp mRNA linear EST 13-SEP-1999
DEFINITION IMAGE:2484240 5', mRNA sequence.
ACCESSION AW021436
VERSION AW021436.1 GI:5874966
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Robertson,N.G., Khetarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R.
and Morton,C.C.
TITLE Isolation of novel and known genes from a human fetal cochlear cDNA
library using subtractive hybridization and differential screening
JOURNAL Genomics 23, 42-50 (1994)
PUBMED 7829101
COMMENT Contact: Morton, C. C.
Departments of Pathology and Obstetrics, Gynecology and
Reproductive Biology
Brigham and Women's Hospital
75 Francis Street, Harvard Medical School, Boston, MA 02115, USA
Tel: 617 732 7980
Fax: 617 738 6996
Email: cmorton@rics.bwh.harvard.edu
DNA sequencing and analyses were performed by National Institutes
of Health Intramural Sequencing Center (NISC; see
http://www.nisc.nih.gov).
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Plate: LLAM6181 row: L column: 1
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1. .372
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2484240"
/tissue_type="cochlea"
/dev_stage="16-22 week fetus"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="Morton Fetal Cochlea"
/note="Organ: ear; Vector: pBluescript SK-; Site 1: ECoRI;
Site 2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned
unidirectionally. Primer: Oligo dt. Fetal cochlea, normal.
37% of inserts <0.5 Kb, 56% 0.5-1.0 Kb, 7% >1 Kb. Uni-ZAP
XR Vector. Library constructed by N. Robertson, C. Morton.
-5' adaptor sequence: 5' GAATTCGGCAGCG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTTTTTTTTTTT 3'"

ORIGIN
Alignment Scores:
Pred. No.: 6.51e-05 Length: 372
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW021436 (1-372)

```

```

QY 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
Db 257 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 310

RESULT 8
LOCUS AW004664/c
DEFINITION 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.

ACCESSION AI758588
VERSION 1
KEYWORDS 387 bp mRNA linear EST 16-DEC-1999
SOURCE cy07h10.x1 NCI CGAP Ut3 Homo sapiens cDNA clone IMAGE:2278435 3'
ORGANISM similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.

ACCESSION AI758588.1 GI:5152313
VERSION 1
KEYWORDS 387 bp mRNA linear EST 16-DEC-1999
SOURCE cy07h10.x1 NCI CGAP Ut3 Homo sapiens cDNA clone IMAGE:2278435 3'
ORGANISM similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.

REFERENCE 1 (bases 1 to 387)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1375 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 217.
Location/Qualifiers
1..387
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2278435"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Ut3"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: Salt;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

ORIGIN
Alignment Scores:
Pred. No.: 6,81e-05 Length: 387
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AI758588 (1-387)

QY 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
Db 91 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 38

RESULT 9
LOCUS AW004664/c
DEFINITION 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.

ACCESSION AW004664
VERSION 1
KEYWORDS 393 bp mRNA linear EST 08-MAR-2000
SOURCE ws91f04.x1 NCI CGAP Co3 Homo sapiens cDNA clone IMAGE:2505343 3'
ORGANISM similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.

```

```

VERSION AW004664.1 GI:5853532
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 393)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 556 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 332.
Location/Qualifiers
1..393
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2505343"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/clone_lib="NCI CGAP Co3"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p7T3 vector. Library went through one round of
normalization."

ORIGIN
Alignment Scores:
Pred. No.: 6,93e-05 Length: 393
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AW004664 (1-393)

QY 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
Db 93 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTTCAGCTA 40

RESULT 10
LOCUS AA287599/c
DEFINITION 1 ThrSerArgLeuArgSerHisIleAnProThrGlyThrValLeuLeuGlnLeu 18
similar to gb: X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);
mRNA sequence.
ACCESSION AA287599
VERSION 1
KEYWORDS 397 bp mRNA linear EST 15-AUG-1997
SOURCE AA287599.1 GI:1933299
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 397)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2004 Std Error: 0.00
Seq primer: -41ml3 fwd. Et from Amersham
High quality sequence stop: 381.
Location/Qualifiers
1. 397
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:5740423"
/db_xref="taxon:9606"
/clone="IMAGE:700989"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP GCBI"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Alignment Scores:
Pred. No.: 7.01e-05 Length: 397
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservat: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-602-220-29 (1-18) x AA287599 (1-397)
QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
DB 100 ACTTCACAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCCTTCAGCTA 47
RESULT 11
AW449396/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AW449396
UI-H-B13-aki-b-05-0-UI.s1 NCI CGAP_Sub5 Homo sapiens cDNA clone
IMAGE:2734688 3', mRNA sequence.
AW449396
AW449396.1 GI:6990172
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 398)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
NCI-CGAP clone distribution information can be found through the
I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1. 398
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2734688"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP Sub5"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub5
is a subtracted library derived from NCI CGAP_Sub4. The
NCI CGAP Sub5 library had 3 million recombinants. A
single-stranded DNA preparation of NCI CGAP Sub4 was used
as a tracer in a subtractive hybridization with a driver
comprising the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)
1323376-1323911, 1456008-1456775, 1500552-1502855);
NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
(IMAGE Clones) 1323912-1325831, 1471368-1472903,
1492104-1493255); NCI CGAP_Lu5 pool 1 LLAM 3575-3582,
3851-3854 (IMAGE Clones)
1414920-1417991, 1520904-1522439); NCI CGAP_GC4 pool 1 LLAM
3164-3167, 3716-3720, 3733-3735 (IMAGE Clones)
1257096-1258631, 1469064-1470983, 1475592-1476743);
NCI CGAP Pr-22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
(IMAGE Clones) 985608-986759, 1101192-1101959,
1217928-1220615); NCI CGAP_Col0 pool 1 LLAM 2644-2653,
2871-2872 (IMAGE Clones)
1057416-1061255, 114584-1145351). (10% of the driver
population), plus a pool of 3,840 arrayed clones from
NCI CGAP Sub1 (IMAGE Clones) 2708616-2710535) and
NCI CGAP Sub2 (IMAGE Clones) 2710536-2712455) (10% of
the driver population), plus a pool of 11,136 clones from
NCI CGAP Sub3 (IMAGE Clones) 2712456-2723591) (10% of the
driver population), plus a pool of 5,472 clones from
NCI CGAP Sub4 (IMAGE Clones) 2723592-2728969) (70% of the
driver population). Subtraction was performed as
previously described [Bonaldo, Lennon & Soares (1996):
Normalization and Subtraction: Two Approaches To
Facilitate Gene Discovery. Genome Research 6, 791-806.
TAG_TISSUE=kidney
TAG_LIB=NCI CGAP_Kids
TAG_SEQ=ATTC"
ORIGIN
Alignment Scores:
Pred. No.: 7.03e-05 Length: 398
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservat: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-602-220-29 (1-18) x AW449396 (1-398)
QY 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
DB 108 ACTTCACAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCCTTCAGCTA 55
RESULT 12
AW448576/c
LOCUS
DEFINITION
ACCESSION
1948576
wq07b03.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2470541 3',
similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);,
mRNA sequence.
1948576
ACCESSION

```

VERSION A1948576.1 GI:5740886
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 401)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 466 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES
 Location/Qualifiers
 1..401
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2470541"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Kid12"
 /note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Alignment Scores:
 Pred. No.: 7,098-05 Length: 401
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x A1948576 (1-401)
 Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
 Db 94 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTGCTTCAGCTA 41
 |||||
 |||||

RESULT 13
 A1949177/c
 LOCUS A1949177 403 bp mRNA linear EST 08-MAR-2000
 DEFINITION wq08h03.x1 NCI CGAP Kid12 Homo sapiens cDNA clone IMAGE:2470709 3', similar to gb:U14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);, mRNA sequence.
 ACCESSION A1949177
 VERSION A1949177.1 GI:5741487
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 403)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 496 Std Error: 0.00
 Seq primer: -40UP from Gibco.

FEATURES
 Location/Qualifiers
 1..403
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2470709"
 /tissue_type="2 pooled tumors (clear cell type)"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Kid12"
 /note="Organ: Kidney; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP Kid5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1323912-1325831, 1471368-1472903 and 1492104-1493255). Subtraction by Bento Soares and M. Fatima Bonaldo."

ORIGIN
 Alignment Scores:
 Pred. No.: 7,138-05 Length: 403
 Score: 89.00 Matches: 18
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x A1949177 (1-403)
 Oy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
 Db 96 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTGCTTCAGCTA 43
 |||||
 |||||

RESULT 14
 AAS48321/c
 LOCUS AAS48321 427 bp mRNA linear EST 05-SEP-1997
 DEFINITION nk16e01.61 NCI CGAP Coll Homo sapiens cDNA clone IMAGE:1013688 3', similar to gb:X14448_rnal ALPHA-GALACTOSIDASE A PRECURSOR (HUMAN);, mRNA sequence.
 ACCESSION AAS48321
 VERSION AAS48321.1 GI:2318603
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 427)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Elias Campo, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Prepared by: Greg Lennon, Ph.D.
cDNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1118 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 364.

FEATURES
source

1. 427
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1013688"
/tissue type="tumor"
/lab host="SOLR (kanamycin resistant)"
/clone lib="NCI CGAP Coll"
/note="Organ: colon; Vector: Bluescript SK-; Site:1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Multiple colon tumors. 5' adaptor sequence: 5'
GAATTCGGCAGCAG 3', 3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

ORIGIN

Alignment Scores:
Pred. No.: 7.61e-05 Length: 427
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AA548321 (1-427)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 90 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTA 37

RESULT 15
AUI49911/c
LOCUS AUI49911 NT2RP2 Homo sapiens cDNA clone NT2RP2000908 3', mRNA
DEFINITION AUI49911 437 bp mRNA linear EST 05-AUG-2002
sequence.
ACCESSION AUI49911
VERSION AUI49911.1 GI:11011432
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 437)
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Yanamoto, J., Wakamatsu, A., Ozawa, M., Nakamura, Y., Nagai, T.,
Sugano, S. and Isogai, T.).
TITLE HRI human cDNA project (Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S.,
Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Ozawa, M.,
Nakamura, Y., Nagai, T., Sugano, S., Isogai, T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5', - & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES
source

Location/Qualifiers
1. 437
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP2000908"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone lib="NT2RP2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"

ORIGIN

Alignment Scores:
Pred. No.: 7.82e-05 Length: 437
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-602-220-29 (1-18) x AUI49911 (1-437)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 91 ACTTCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTA 38

Search completed: January 1, 2006, 05:47:06
Job time : 330.524 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:29:37 ; Search time 14.6483 Seconds
(without alignments)
2184.292 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHNPRTGTVLLQL 18

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140559_5451/app_query.fasta_1.782
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=patn.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blowsum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1 1 237 @runat_30122005_140559_5451 -NCFU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PPCTUS COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	210	3	US-08-928-881-19
2	89	100.0	210	3	US-09-543-921-19
3	89	100.0	210	3	US-09-266-014-1
4	89	100.0	210	3	US-09-491-759-19
5	89	100.0	297	2	US-07-602-824A-8
6	89	100.0	297	2	US-07-983-451-8
7	89	100.0	297	2	US-08-261-577-11
8	89	100.0	1158	3	US-09-176-666-53
9	89	100.0	1161	3	US-09-176-666-52

10	89	100.0	1164	3	US-09-176-666-51	Sequence 51, Appl
11	89	100.0	1167	3	US-09-176-666-50	Sequence 50, Appl
12	89	100.0	1170	3	US-09-176-666-49	Sequence 49, Appl
13	89	100.0	1173	3	US-09-176-666-48	Sequence 48, Appl
14	89	100.0	1176	3	US-09-176-666-47	Sequence 47, Appl
15	89	100.0	1179	3	US-09-176-666-46	Sequence 46, Appl
16	89	100.0	1182	3	US-09-176-666-45	Sequence 45, Appl
17	89	100.0	1188	3	US-09-176-666-44	Sequence 44, Appl
18	89	100.0	1194	3	US-09-176-666-43	Sequence 43, Appl
19	89	100.0	1197	3	US-08-928-881-25	Sequence 25, Appl
20	89	100.0	1197	3	US-09-543-921-25	Sequence 25, Appl
21	89	100.0	1197	3	US-09-266-014-25	Sequence 25, Appl
22	89	100.0	1197	3	US-09-491-759-25	Sequence 25, Appl
23	89	100.0	1214	3	US-09-023-655-1001	Sequence 1001, Ap
24	89	100.0	1254	3	US-09-993-059-15	Sequence 15, Appl
25	89	100.0	1254	3	US-10-103-327-15	Sequence 15, Appl
26	89	100.0	1266	3	US-09-993-059-11	Sequence 11, Appl
27	89	100.0	1266	3	US-10-103-327-11	Sequence 11, Appl
28	89	100.0	1272	3	US-09-993-059-17	Sequence 17, Appl
29	89	100.0	1278	3	US-10-103-327-17	Sequence 17, Appl
30	89	100.0	1278	3	US-09-993-059-7	Sequence 7, Appl
31	89	100.0	1278	3	US-10-103-327-7	Sequence 7, Appl
32	89	100.0	1284	3	US-09-993-059-13	Sequence 13, Appl
33	89	100.0	1284	3	US-10-103-327-13	Sequence 13, Appl
34	89	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
35	89	100.0	1296	3	US-10-103-327-3	Sequence 3, Appl
36	89	100.0	1296	3	US-09-993-059-9	Sequence 9, Appl
37	89	100.0	1296	3	US-10-103-327-9	Sequence 9, Appl
38	89	100.0	1304	9	5179023-3	Patent No. 5179023
39	89	100.0	1308	3	US-09-993-059-5	Sequence 5, Appl
40	89	100.0	1308	3	US-10-103-327-5	Sequence 5, Appl
41	89	100.0	1343	3	US-08-928-881-18	Sequence 18, Appl
42	89	100.0	1343	3	US-09-543-921-18	Sequence 18, Appl
43	89	100.0	1343	3	US-09-266-014-3	Sequence 3, Appl
44	89	100.0	1343	3	US-09-491-759-18	Sequence 18, Appl
45	89	100.0	1393	2	US-07-602-824A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-928-881-19
; Sequence 19, Application US/08928881
; Patent No. 6083725
; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; TITLE OF INVENTION: DEFICIENCY
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,881
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frazer, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 07236/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:
LENGTH: 210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-928-881-19

Alignment Scores:
Pred. No.: 2.06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-08-928-881-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACGTGTTTGGCTTCAGCTA 208

RESULT 2

US-09-543-921-19
Sequence 19, Application US/09543921
Patent No. 6395884
GENERAL INFORMATION:
APPLICANT: Selden et al., Richard F.
TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A DEFICIENCY

NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/543,921

FILING DATE: 06-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,881

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 07236/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 210 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-543-921-19

Alignment Scores:
Pred. No.: 2.06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-543-921-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACGTGTTTGGCTTCAGCTA 208

RESULT 3

US-09-266-014-1
Sequence 1, Application US/09266014
Patent No. 6458574

GENERAL INFORMATION:

APPLICANT: Selden, Richard F.

APPLICANT: Borowski, Marianne

APPLICANT: Kinoshita, Carol M

APPLICANT: Treco, Douglas A

APPLICANT: Williams, Melanie D

APPLICANT: Schuetz, Thomas J

APPLICANT: Daniel, Peter F.

TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

FILE REFERENCE: FABRY DISEASE (18082-001)

CURRENT APPLICATION NUMBER: US/09/266,014

CURRENT FILING DATE: 1999-03-11

PRIOR APPLICATION NUMBER: 60/026,041

PRIOR FILING DATE: 1996-09-13

PRIOR APPLICATION NUMBER: 08/928,881

PRIOR FILING DATE: 1997-09-12

PRIOR APPLICATION NUMBER: PCT US97/16603

PRIOR FILING DATE: 1997-09-12

NUMBER OF SEQ ID NOS: 24

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1

LENGTH: 210

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Human

OTHER INFORMATION: fibroblast library probe: exon 7, including

OTHER INFORMATION: amplification primers.

US-09-266-014-1

Alignment Scores:

Pred. No.: 2.06e-07 Length: 210

Score: 89.00 Matches: 18

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-266-014-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAGTCACATAAATCCACAGGCACGTGTTTGGCTTCAGCTA 208

RESULT 4

US-09-491-759-19
Sequence 19, Application US/09491759
Patent No. 6566099

GENERAL INFORMATION:

APPLICANT: Selden et al., Richard F.

TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A

DEFICIENCY

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/491,759
FILING DATE: 27-Jan-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/928,881
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fraser, Janis K.
REGISTRATION NUMBER: 34,819
REFERENCE/DOCKET NUMBER: 07236/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
SEQUENCE CHARACTERISTICS:
LENGTH: 210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-09-491-759-19
Alignment Scores:
Pred. No.: 2,06e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-491-759-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleasnProThrGlyThrValLeuLeuGlnLeu 18
Db 155 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCGTTTCAGCTA 208

RESULT 5
US-07-602-824A-8
Sequence 8, Application US/07602824A
Patent No. 5356804
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David P.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
TITLE OF INVENTION: Active alpha-GALACTOSIDASE A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/602,824A
FILING DATE: 24-OCT-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..279
US-07-602-824A-8

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-07-602-824A-8 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleasnProThrGlyThrValLeuLeuGlnLeu 18
Db 7 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCGTTTCAGCTA 60

RESULT 6
US-07-983-451-8
Sequence 8, Application US/07983451
Patent No. 5401650
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David P.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-Galactosidase A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/983,451
FILING DATE: 30-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 07/983,451
REFERENCE/DOCKET NUMBER: 6923-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..279
US-07-983-451-8

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-07-983-451-8 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 7 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTTCAGCTA 60

RESULT 7

US-08-261-577-11
; Sequence 11, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994

CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 297 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..279

US-08-261-577-11

Alignment Scores:
Pred. No.: 3.16e-07 Length: 297
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 2 Gaps: 0

US-10-602-220-29 (1-18) x US-08-261-577-11 (1-297)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18

Db 7 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTTCAGCTA 60

RESULT 8

US-09-176-666-53
; Sequence 53, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 1158
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-53

Alignment Scores:
Pred. No.: 1.69e-06 Length: 1158
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-53 (1-1158)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTTCAGCTA 1158

RESULT 9

US-09-176-666-52
; Sequence 52, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FABRY DISEASE
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 1161
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-52

Alignment Scores:
Pred. No.: 1.69e-06 Length: 1161
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-52 (1-1161)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGAACTGTTTTCAGCTA 1158

```
RESULT 10
US-09-176-666-51
; Sequence 51, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 1164
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-51

Alignment Scores:
Pred. No.: 1.7e-06 Length: 1164
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-51 (1-1164)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 11
US-09-176-666-50
; Sequence 50, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-50

Alignment Scores:
Pred. No.: 1.7e-06 Length: 1167
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-50 (1-1167)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 12
US-09-176-666-49
; Sequence 49, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 49
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-49

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1170
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-49 (1-1170)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 13
US-09-176-666-48
; Sequence 48, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-48

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1173
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-48 (1-1173)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 14
US-09-176-666-47
; Sequence 47, Application US/09176666
; Patent No. 6210666
; GENERAL INFORMATION:
; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; EARLIER FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-47

Alignment Scores:
Pred. No.: 1.71e-06 Length: 1173
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-47 (1-1173)
Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158
```

; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FERRY DISEASE

; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 47
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-47

Alignment Scores:
Pred. No.: 1.72e-06 Length: 1176
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-47 (1-1176)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

RESULT 15

US-09-176-666-46
; Sequence 46, Application US/09176666
; Patent No. 6210666

; GENERAL INFORMATION:

; APPLICANT: Miyamura, No. 6210666uhiro
; TITLE OF INVENTION: TRUNCATED alpha-GALACTOSIDASE A TO TREAT
; TITLE OF INVENTION: FERRY DISEASE

; FILE REFERENCE: 101.018US1
; CURRENT APPLICATION NUMBER: US/09/176,666
; CURRENT FILING DATE: 1998-10-21
; EARLIER APPLICATION NUMBER: 60/062,650
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 46
; LENGTH: 1179
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-176-666-46

Alignment Scores:
Pred. No.: 1.73e-06 Length: 1179
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-176-666-46 (1-1179)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAGGTTAAGAAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1158

Search completed: January 1, 2006, 05:53:06
Job time : 15.6483 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 03:35:43 ; Search time 64.4276 Seconds
(without alignments)
2310.327 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10602220/runat_30122005_140601_5538/app_query.fasta_1.782
-DB=Published_Applications_NA_Main -QFWT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602220 CGCN 1.1.1364 @runat_30122005_140601_5538 -NCFU=6 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA_Main:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	210	5	US-10-165-060-1
2	89	100.0	210	5	US-10-165-968-1
3	89	100.0	210	6	US-10-318-905-19
4	89	100.0	227	7	US-10-242-535A-29035
5	89	100.0	227	7	US-10-085-783A-29035
6	89	100.0	270	7	US-10-242-535A-30531
7	89	100.0	270	7	US-10-085-783A-30531

ALIGNMENTS

RESULT 1
US-10-165-060-1
; Sequence 1, Application US/10165060
; Publication No. US20030077806A1
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FABRY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,060
; CURRENT FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: Human
; OTHER INFORMATION: fibroblast library probe: exon 7, including
; OTHER INFORMATION: amplification primers.
US-10-165-060-1

Alignment Scores:
Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-060-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 2

US-10-165-968-1
; Sequence 1, Application US/10165968
; Publication No. US20030113894A1
; GENERAL INFORMATION:
; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.
; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency
; FILE REFERENCE: FARY DISEASE (18082-001)
; CURRENT APPLICATION NUMBER: US/10/165,968
; CURRENT FILING DATE: 2002-06-10
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: PCT US97/16603
; PRIOR FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 1
; LENGTH: 210
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Human
; OTHER INFORMATION: fibroblast library probe: exon 7, including
; OTHER INFORMATION: amplification primers.
US-10-165-968-1

Alignment Scores:
Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-968-1 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 3

US-10-318-905-19
; Sequence 19, Application US/10318905
; Publication No. US20030152560A1

; GENERAL INFORMATION:
; APPLICANT: Selden et al., Richard F.
; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A
; DEFICIENCY

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/318,905

FILING DATE: 12-Dec-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/928,881

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 07236/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 210 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 19:

US-10-318-905-19

Alignment Scores:

Pred. No.: 2.49e-07 Length: 210
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-318-905-19 (1-210)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
|||||
Db 155 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGCTTCAGCTA 208

RESULT 4

US-10-242-535A-29035

; Sequence 29035, Application US/10242535A

; Publication No. US20040013663A1

; GENERAL INFORMATION:

; APPLICANT: ChondroGene Inc.

; APPLICANT: Liew, C.C.

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2005

; CURRENT APPLICATION NUMBER: US/10/242,535A

; CURRENT FILING DATE: 2002-09-12

; PRIOR APPLICATION NUMBER: US 10/085,783

; PRIOR FILING DATE: 2002-02-28

; PRIOR APPLICATION NUMBER: US 60/305,340

; PRIOR FILING DATE: 2001-07-13

; PRIOR APPLICATION NUMBER: US 60/275,017

; PRIOR FILING DATE: 2001-03-12

; PRIOR APPLICATION NUMBER: US 60/271,955

; PRIOR FILING DATE: 2001-02-28

; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29035
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-29035

Alignment Scores:
Pred. No.: 2.72e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-242-535A-29035 (1-227)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 197

RESULT 5

US-10-085-783A-29035
; Sequence 29035, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis

; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29035
; LENGTH: 227
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-29035

Alignment Scores:
Pred. No.: 2.72e-07 Length: 227
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-085-783A-29035 (1-227)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 144 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 197

RESULT 6

US-10-242-535A-30531
; Sequence 30531, Application US/10242535A
; Publication No. US20040013663A1

; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30531
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-30531

Alignment Scores:
Pred. No.: 3.34e-07 Length: 270
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-242-535A-30531 (1-270)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGGCTTCAGCTA 227

RESULT 7

US-10-085-783A-30531
; Sequence 30531, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.

; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30531
; LENGTH: 270
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-30531

Alignment Scores:
Pred. No.: 3.34e-07 Length: 270
Score: 89.00 Matches: 18

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-085-783A-30531 (1-270)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 174 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 227

RESULT 8

US-10-165-060-5
; Sequence 5, Application US/10165060
; Publication No. US20030077806A1

GENERAL INFORMATION:

; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/10/165,060

; CURRENT FILING DATE: 2002-06-07

; PRIOR APPLICATION NUMBER: US/09/266,014

; PRIOR FILING DATE: 1999-03-11

; PRIOR APPLICATION NUMBER: 60/026,041

; PRIOR FILING DATE: 1996-09-13

; PRIOR APPLICATION NUMBER: 08/928,881

; PRIOR FILING DATE: 1997-09-12

; PRIOR APPLICATION NUMBER: PCT US97/16603

; PRIOR FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-165-060-5

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-060-5 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 9

US-10-165-968-5

; Sequence 5, Application US/10165968

; Publication No. US20030113894A1

GENERAL INFORMATION:

; APPLICANT: Selden, Richard F
; APPLICANT: Borowski, Marianne
; APPLICANT: Kinoshita, Carol M
; APPLICANT: Treco, Douglas A
; APPLICANT: Williams, Melanie D
; APPLICANT: Schuetz, Thomas J
; APPLICANT: Daniel, Peter F.

; TITLE OF INVENTION: Treatment for alpha-Galactosidase A Deficiency

; FILE REFERENCE: FABRY DISEASE (18082-001)

; CURRENT APPLICATION NUMBER: US/10/165,968

; CURRENT FILING DATE: 2002-06-10

; PRIOR APPLICATION NUMBER: US/09/266,014
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: 60/026,041
; PRIOR FILING DATE: 1996-09-13
; PRIOR APPLICATION NUMBER: 08/928,881
; PRIOR FILING DATE: 1997-09-12

; PRIOR APPLICATION NUMBER: PCT US97/16603

; PRIOR FILING DATE: 1997-09-12

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 1197

; TYPE: DNA

; ORGANISM: Homo sapiens

US-10-165-968-5

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-165-968-5 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1158

RESULT 10

US-10-318-905-25

; Sequence 25, Application US/10318905

; Publication No. US20030152560A1

GENERAL INFORMATION:

; APPLICANT: Selden et al., Richard F.

; TITLE OF INVENTION: THERAPY FOR ALPHA-GALACTOSIDASE A

; DEFICIENCY

; NUMBER OF SEQUENCES: 28

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/318,905

; FILING DATE: 12-Dec-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/928,881

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Fraser, Janis K.

; REGISTRATION NUMBER: 34,819

; REFERENCE/DOCKET NUMBER: 07236/003001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; INFORMATION FOR SEQ ID NO: 25:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1197 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 25:

US-10-318-905-25

Alignment Scores:
Pred. No.: 1.91e-06 Length: 1197
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-318-905-25 (1-1197)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1105 ACTTCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTTCAGCTA 1158

RESULT 11

US-10-641-643-1001
; Sequence 1001, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; Suffer, J. Stuart
; Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
; GENE EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESS: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/641,643
; FILING DATE: 14-Aug-2003
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: <Unknown>
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1001:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1214 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g178245
; SEQUENCE DESCRIPTION: SEQ ID NO: 1001 :

Alignment Scores:
Pred. No.: 1.94e-06 Length: 1214
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-641-643-1001 (1-1214)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1122 ACTTCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTTCAGCTA 1175

RESULT 12

US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 2.02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-29 (1-18) x US-09-993-059-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
Db 1198 ACTTCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTTCAGCTA 1251

RESULT 13

US-10-103-327-15
; Sequence 15, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15

Alignment Scores:
Pred. No.: 2.02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-10-602-220-29 (1-18) x US-10-103-327-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCAGCTA 1251

RESULT 14

US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15

Alignment Scores:
Pred. No.: 2,02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-602-219-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCAGCTA 1251

RESULT 15

US-10-602-220-15
; Sequence 15, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation

; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-15

Alignment Scores:
Pred. No.: 2,02e-06 Length: 1254
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-29 (1-18) x US-10-602-220-15 (1-1254)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuGlnLeu 18
|||||
Db 1198 ACTTCAAGGTTAAGAAGTCACATAATCCACAGGCACCTGTTTTCAGCTA 1251

Search completed: January 1, 2006, 09:34:34
Job time : 66.4276 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 05:47:17 ; Search time 17.5034 Seconds
(without alignments)
539.100 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSLRSHINPTGVLLQL 18

Scoring table:

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool/US10602220/runat.30122005.140602.5569/app_query.fasta.1.782
-DB=Published Applications NA New -GPMT=fastcap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-NCPLEN=200000000 -USRP=US10602220 @CGN 1.1.233 @runat.30122005.140602.5569
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA New.*

- 1: /cgn2_6/prodata/1/pubpna/US08 NEW PUB.seq.*
- 2: /cgn2_6/prodata/1/pubpna/US06 NEW PUB.seq.*
- 3: /cgn2_6/prodata/1/pubpna/US07 NEW PUB.seq.*
- 4: /cgn2_6/prodata/1/pubpna/US09 NEW PUB.seq.*
- 5: /cgn2_6/prodata/1/pubpna/US10 NEW PUB.seq.*
- 6: /cgn2_6/prodata/1/pubpna/US11 NEW PUB.seq.*
- 7: /cgn2_6/prodata/1/pubpna/US11 NEW PUB.seq.*
- 8: /cgn2_6/prodata/1/pubpna/US11 NEW PUB.seq.*
- 9: /cgn2_6/prodata/1/pubpna/US11 NEW PUB.seq.*
- 10: /cgn2_6/prodata/1/pubpna/US60 NEW PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	1355	US-10-981-267-26	Sequence 26, Appl
2	89	100.0	1547	US-10-981-267-25	Sequence 25, Appl
3	48	53.9	1716	US-10-750-185-42165	Sequence 42165, A
C 4	47	52.8	3321	US-10-750-185-61475	Sequence 61475, A
5	44	49.4	1380	US-10-750-185-56674	Sequence 56674, A
C 6	44	49.4	1794	US-11-074-178-257	Sequence 257, App
7	43	48.3	1587	US-10-131-826A-465	Sequence 465, App
8	43	48.3	1710	US-10-750-185-24794	Sequence 24794, A

C 9	43	48.3	1710	6	US-10-750-185-50667	Sequence 50667, A
10	43	48.3	1861	6	US-10-750-185-43369	Sequence 43369, A
11	43	48.3	100001	6	US-10-944-272-1	Sequence 1, Appl
12	43	48.3	207835	7	US-11-121-086-39	Sequence 39, Appl
13	43	48.3	207835	7	US-11-121-086-40	Sequence 40, Appl
14	42	47.2	647	6	US-10-816-768-99	Sequence 99, Appl
15	42	47.2	1916	6	US-10-750-185-38650	Sequence 38650, A
C 16	42	47.2	2217	6	US-10-750-185-40730	Sequence 40730, A
17	42	47.2	2462	6	US-10-750-185-44108	Sequence 44108, A
18	42	47.2	6855	6	US-10-750-185-32086	Sequence 32086, A
C 19	42	47.2	10129	7	US-11-044-111-21	Sequence 21, Appl
C 20	42	47.2	31100	6	US-10-829-826B-24	Sequence 24, Appl
C 21	42	47.2	49979	6	US-10-995-561-13443	Sequence 13443, A
C 22	42	47.2	184000	7	US-11-121-086-37	Sequence 37, Appl
23	41	46.1	25	7	US-11-121-849-5271128	Sequence 527128, A
24	41	46.1	201	6	US-10-995-561-8091	Sequence 8091, Ap
25	41	46.1	201	6	US-10-995-561-8098	Sequence 8098, Ap
26	41	46.1	201	6	US-10-995-561-8132	Sequence 8132, Ap
27	41	46.1	201	6	US-10-995-561-8172	Sequence 8172, Ap
28	41	46.1	201	6	US-10-995-561-8179	Sequence 8179, Ap
29	41	46.1	201	6	US-10-995-561-37193	Sequence 37193, A
C 30	41	46.1	201	6	US-10-995-561-41124	Sequence 41124, A
31	41	46.1	201	6	US-10-995-561-73965	Sequence 73965, A
32	41	46.1	699	6	US-10-750-185-25901	Sequence 25901, A
C 33	41	46.1	826	6	US-10-750-185-38506	Sequence 38506, A
34	41	46.1	1156	6	US-10-750-185-45065	Sequence 45065, A
C 35	41	46.1	1201	6	US-10-750-185-27310	Sequence 27310, A
C 36	41	46.1	1883	6	US-10-750-185-47129	Sequence 47129, A
37	41	46.1	2606	6	US-10-750-185-49768	Sequence 49768, A
38	41	46.1	2703	7	US-11-191-072-1	Sequence 1, Appl
C 39	41	46.1	3885	6	US-10-793-626-3883	Sequence 3883, Ap
40	41	46.1	4250	6	US-10-995-561-263	Sequence 263, App
41	41	46.1	4432	6	US-10-995-561-265	Sequence 265, App
42	41	46.1	4511	6	US-10-995-561-264	Sequence 264, App
43	41	46.1	5700	6	US-10-513-786-7	Sequence 7, Appl
44	41	46.1	5700	6	US-10-513-786-9	Sequence 9, Appl
45	41	46.1	11115	6	US-10-513-786-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens

US-10-981-267-26

Alignment Scores:

Pred. No.: 3,99e-06 Length: 1355
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-981-267-26 (1-1355)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1245 ACTTCAAGGTAAAGAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1298

RESULT 2

US-10-981-267-25
; Sequence 25, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; TYPE: DNA
; LENGTH: 1547
; FEATURE:
; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence
US-10-981-267-25

Alignment Scores:

Pred. No.: 4.67e-06 Length: 1547
Score: 89.00 Matches: 18
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-981-267-25 (1-1547)

Qy 1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1248 ACTTCAAGGTAAAGAGTCACATAATCCACAGGCACTGTTTTCCTTCAGCTA 1301

RESULT 3

US-10-185-42165
; Sequence 42165, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.

; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42165
; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Bovine 19866881152718
US-10-750-185-42165

Alignment Scores:

Pred. No.: 163 Length: 1716
Score: 48.00 Matches: 8
Percent Similarity: 86.67% Conservative: 5
Best Local Similarity: 53.33% Mismatches: 2
Query Match: 53.93% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-42165 (1-1716)

Qy 4 LeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 1145 CTTCAACAGAACTTGAACTTACTGGAACCTTGATCTTAAGACTT 1189

RESULT 4

US-10-750-185-61475/c
; Sequence 61475, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61475
; LENGTH: 3321
; TYPE: DNA
; ORGANISM: Bovine 19866880527707
US-10-750-185-61475

Alignment Scores:

Pred. No.: 546 Length: 3321
Score: 47.00 Matches: 9
Percent Similarity: 68.75% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 5
Query Match: 52.81% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-61475 (1-3321)

Qy 3 ArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 772 CGGCTCCATTGCCATGATGAGGCCCTCGTGGCTCAGTGTCTTGGCTCTC 725

RESULT 5

US-10-185-185-56674
; Sequence 56674, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56674
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Bovine 19866881332286
US-10-750-185-56674

Alignment Scores:

Pred. No.:	674	Length:	1380
Score:	44.00	Matches:	7
Percent Similarity:	81.25%	Conservative:	6
Best Local Similarity:	43.75%	Mismatches:	3
Query Match:	49.44%	Indels:	0
DB:	6	Gaps:	0

US-10-602-220-29 (1-18) x US-10-750-185-56674 (1-1380)

Qy 2 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 17
Db 1220 AACCGAGTCAAAAGGAACTAAATCCACAGCTACACTTATTTTACAA 1267

RESULT 6

US-11-074-176-257/c
; Sequence 257, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 257
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1794)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: ORF 1763; oligopeptidase
US-11-074-176-257

Alignment Scores:

Pred. No.:	1794	Length:	1794
Score:	44.00	Matches:	9
Percent Similarity:	66.87%	Conservative:	1
Best Local Similarity:	60.00%	Mismatches:	5

Pred. No.:	922	Length:	1794
Score:	44.00	Matches:	8
Percent Similarity:	80.00%	Conservative:	4
Best Local Similarity:	53.33%	Mismatches:	3
Query Match:	49.44%	Indels:	0
DB:	7	Gaps:	0

US-10-602-220-29 (1-18) x US-11-074-176-257 (1-1794)

Qy 2 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeu 16
Db 59 ACACGGTTAAGTCCCAATTTAATCTTCTGGAACCTTCACCTCTT 15

RESULT 7

US-10-131-826A-465
; Sequence 465, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 465
; LENGTH: 1587
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-131-826A-465

Alignment Scores:

Pred. No.:	1.21e+03	Length:	1587
Score:	43.00	Matches:	9
Percent Similarity:	66.87%	Conservative:	1
Best Local Similarity:	60.00%	Mismatches:	5

Alignment Scores:	4.07e+05	Length:	207835
Pred. No.:	43.00	Matches:	7
Score:	69.23%	Conservative:	2
Percent Similarity:	53.85%	Mismatches:	4
Best Local Similarity:	48.31%	Indels:	0
Query Match:	7	Gaps:	0
DB:			
US-10-602-220-29 (1-18) x US-11-121-086-40 (1-207835)			
QY	1 ThrSerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 17		
DB	9195 ACAGGTGAGCTCCGCACTCATGTGGTCCATGTGGAACA 9233		
RESULT 14			
US-10-816-768-99			
Sequence 99, Application US/10816768			
Publication No. US20050250936A1			
GENERAL INFORMATION:			
APPLICANT: Oppermann, Hermann			
APPLICANT: Tai, Mei-Sheng			
APPLICANT: McCartney, John			
TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins			
FILE REFERENCE: STK-075			
CURRENT APPLICATION NUMBER: US/10/816,768			
CURRENT FILING DATE: 2004-04-02			
NUMBER OF SEQ ID NOS: 124			
SOFTWARE: PatentIn version 2.0			
SEQ ID NO 99			
LENGTH: 647			
TYPE: DNA			
ORGANISM: Homo sapiens			
FEATURE:			
OTHER INFORMATION: H2528			
US-10-816-768-99			
Alignment Scores:	630	Length:	647
Pred. No.:	42.00	Matches:	7
Score:	71.43%	Conservative:	3
Percent Similarity:	50.00%	Mismatches:	4
Best Local Similarity:	47.19%	Indels:	0
Query Match:	6	Gaps:	0
DB:			
US-10-602-220-29 (1-18) x US-10-816-768-99 (1-647)			
QY	4 LeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 17		
DB	417 TTGCGTTCCGCACTTCGAGCCACCACCATGCCATCATTCAG 458		
RESULT 15			
US-10-185-38650			
Sequence 38650, Application US/10750185			
Publication No. US20050260603A1			
GENERAL INFORMATION:			
APPLICANT: MMI GENOMICS, INC.			
APPLICANT: Denise, Sue K.			
APPLICANT: Kerr, Richard			
APPLICANT: ROSENFELD, David			
APPLICANT: HOLM, Tom			
APPLICANT: BATES, Stephen			
APPLICANT: FANTIN, Dennis			
TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS			
FILE REFERENCE: MM1100-2			
CURRENT APPLICATION NUMBER: US/10/750,185			
CURRENT FILING DATE: 2003-12-31			
PRIOR APPLICATION NUMBER: US 60/437,482			
PRIOR FILING DATE: 2002-12-31			
NUMBER OF SEQ ID NOS: 64922			
SOFTWARE: PatentIn version 3.1			
SEQ ID NO 38650			
LENGTH: 1916			
TYPE: DNA			

! ORGANISM: Bovine 19866880881467
US-10-750-185-38650

Alignment Scores:
Pred. No.: 2.31e+03 Length: 1916
Score: 42.00 Matches: 9
Percent Similarity: 70.59% Conservative: 3
Best Local Similarity: 52.94% Mismatches: 5
Query Match: 47.19% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-29 (1-18) x US-10-750-185-38650 (1-1916)

Qy 2 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 18
Db 955 TCTATGTTGGCTTCAGGTATAAGCCTAGTGGTTTCAGTATTTTTCAGATT 1005

Search completed: January 1, 2006, 09:56:14
Job time : 58.5034 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:52 ; Search time 0.910345 Seconds
(without alignments)
148.074 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSRLRSHNPTGTVLLQL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 57103 seqs, 7488799 residues

Total number of hits satisfying chosen parameters: 57103

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA_New.*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	47.2	102	6	US-10-816-768-84
2	42	47.2	102	6	US-10-816-768-85
3	42	47.2	102	6	US-10-816-768-86
4	42	47.2	102	6	US-10-816-768-87
5	42	47.2	102	6	US-10-816-768-88
6	42	47.2	203	6	US-10-816-768-100
7	41	46.1	101	7	US-11-191-072-14
8	41	46.1	102	6	US-10-816-768-83
9	41	46.1	119	7	US-11-191-072-12
10	41	46.1	119	7	US-11-191-072-13
11	41	46.1	501	7	US-11-191-072-2
12	41	46.1	1020	6	US-10-513-786-4
13	41	46.1	1041	6	US-10-995-561-780
14	41	46.1	1041	6	US-10-995-561-782
15	41	46.1	1097	6	US-10-995-561-781
16	41	46.1	1900	6	US-10-513-786-3
17	41	46.1	3704	6	US-10-513-786-1
18	38	42.7	218	6	US-10-467-657-1398
19	38	42.7	362	7	US-11-010-874-18
20	38	42.7	418	7	US-11-077-716-2
21	37	41.6	30	6	US-10-467-657-7384
22	37	41.6	402	6	US-10-821-234-1581
23	36	40.4	107	7	US-11-053-076-45
24	36	40.4	235	7	US-11-188-473-2
25	36	40.4	418	6	US-10-821-234-1331

ALIGNMENTS

RESULT 1

US-10-816-768-84
; Sequence 84, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 84
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-2/GDF-6
US-10-816-768-84

Query Match 47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVLLQ 17

Db 38 LRSHLEPTNHAIIQ 51

RESULT 2

US-10-816-768-85
; Sequence 85, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 85
; LENGTH: 102
; TYPE: PRT

```
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: GDF-6
US-10-816-768-85

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 3
US-10-816-768-86
; Sequence 86, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 86
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: CDMF-2
US-10-816-768-86

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 4
US-10-816-768-87
; Sequence 87, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 87
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: GDF-7
US-10-816-768-87

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||

; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: GDF-6
US-10-816-768-85

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 5
US-10-816-768-88
; Sequence 88, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 88
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-3 construct
US-10-816-768-88

Query Match      47.2%; Score 42; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 0.83;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      38 LRSHLEPTNHAIQ 51

RESULT 6
US-10-816-768-100
; Sequence 100, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 2.0
; SEQ ID NO 100
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: H2528
US-10-816-768-100

Query Match      47.2%; Score 42; DB 6; Length 203;
Best Local Similarity 50.0%; Pred. No. 1.8;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy      4 LRSHNPTGTVLLQ 17
      ||||: ||: ||
Db      139 LRSHLEPTNHAIQ 152

RESULT 7
US-11-191-072-14
; Sequence 14, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
```


; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; PRIOR FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 101
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (66)..(66)
; OTHER INFORMATION: Xaa = any amino acid except cysteine
US-11-191-072-14

Query Match 46.1%; Score 41; DB 7; Length 101;
Best Local Similarity 50.0%; Pred. No. 1.2; Mismatches 4; Indels 0; Gaps 0;
Matches 7; Conservative 3;

QY 4 LRSHNPTGTVLLQ 17
||||: || :
Db 38 LRSHLEPTNHAVIQ 51

RESULT 8
US-10-816-768-83
; Sequence 83, Application US/10816768
; Publication No. US20050250936A1
; GENERAL INFORMATION:
; APPLICANT: Oppermann, Hermann
; APPLICANT: Tai, Mei-Sheng
; APPLICANT: McCartney, John
; TITLE OF INVENTION: Modified TGF-beta Superfamily Proteins
; FILE REFERENCE: STK-075
; CURRENT APPLICATION NUMBER: US/10/816,768
; CURRENT FILING DATE: 2004-04-02
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn version 2.0
; SEQ ID NO 83
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CDMF-1/GDF-5
US-10-816-768-83

Query Match 46.1%; Score 41; DB 6; Length 102;
Best Local Similarity 50.0%; Pred. No. 1.2; Mismatches 4; Indels 0; Gaps 0;
Matches 7; Conservative 3;

QY 4 LRSHNPTGTVLLQ 17
||||: || :
Db 38 LRSHLEPTNHAVIQ 51

RESULT 9
US-11-191-072-12
; Sequence 12, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072

; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: mutated recombinant human MP52
US-11-191-072-12

Query Match 46.1%; Score 41; DB 7; Length 119;
Best Local Similarity 50.0%; Pred. No. 1.5; Mismatches 4; Indels 0; Gaps 0;
Matches 7; Conservative 3;

QY 4 LRSHNPTGTVLLQ 17
||||: || :
Db 55 LRSHLEPTNHAVIQ 68

RESULT 10
US-11-191-072-13
; Sequence 13, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 119
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: monomeric mutated recombinant human MP52
; NAME/KEY: MISC FEATURE
; LOCATION: (83)..(83)
; OTHER INFORMATION: Xaa = any amino acid except cysteine
US-11-191-072-13

Query Match 46.1%; Score 41; DB 7; Length 119;
Best Local Similarity 50.0%; Pred. No. 1.5; Mismatches 4; Indels 0; Gaps 0;
Matches 7; Conservative 3;

QY 4 LRSHNPTGTVLLQ 17
||||: || :
Db 55 LRSHLEPTNHAVIQ 68

RESULT 11
US-11-191-072-2
; Sequence 2, Application US/11191072
; Publication No. US20050282255A1
; GENERAL INFORMATION:
; APPLICANT: Hotten, Gertrud

```
; APPLICANT: Bechtold, Rolf
; APPLICANT: Pohl, Jens
; TITLE OF INVENTION: Monomeric Protein of the TGF-beta Family
; FILE REFERENCE: 2923-128
; CURRENT APPLICATION NUMBER: US/11/191,072
; CURRENT FILING DATE: 2005-07-28
; PRIOR APPLICATION NUMBER: US/10/048,458
; PRIOR FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: PCT/EP00/07600
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: EP 99115613.4
; PRIOR FILING DATE: 1999-08-06
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (465)..(465)
; OTHER INFORMATION: Xaa = any amino acid
US-11-191-072-2

Query Match          46.1%; Score 41; DB 7; Length 501;
Best Local Similarity 50.0%; Pred. No. 7.7;
Matches 7; Conservative 3; Mismatches 4; Indels 4; Gaps 0;

Qy      4 LRSHNPQTGVLLQ 17
Db      437 LRSHLEPTNHAIVQ 450

RESULT 12
US-10-513-786-4
; Sequence 4, Application US/10513786
; Publication No. US20050260589A1
; GENERAL INFORMATION:
; APPLICANT: Bio Control Institute Limited
; APPLICANT: OHTA, Michio
; APPLICANT: AGATA, Norio
; TITLE OF INVENTION: A cereulide synthetase, a gene thereof, and a detection method fo
; TITLE OF INVENTION: cereulide.
; FILE REFERENCE: P0202401
; CURRENT APPLICATION NUMBER: US/10/513,786
; CURRENT FILING DATE: 2004-11-16
; PRIOR APPLICATION NUMBER: JP P2002-142398
; PRIOR FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1020
; TYPE: PRT
; ORGANISM: Bacillus cereus
US-10-513-786-4

Query Match          46.1%; Score 41; DB 6; Length 1020;
Best Local Similarity 57.9%; Pred. No. 17;
Matches 11; Conservative 2; Mismatches 4; Indels 2; Gaps 1;

Qy      1 TSLRSHI--NPTGTVLLQ 17
Db      885 TLALRSHLDNPTVDVLLK 903

RESULT 13
US-10-995-561-780
; Sequence 780, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 781
; LENGTH: 1097
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-781

Query Match          46.1%; Score 41; DB 6; Length 1097;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 780
; LENGTH: 1041
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-780

Query Match          46.1%; Score 41; DB 6; Length 1041;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      5 RSHINPTGTV 14
Db      808 RQHLEPTGTI 817

RESULT 14
US-10-995-561-782
; Sequence 782, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 782
; LENGTH: 1041
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-782

Query Match          46.1%; Score 41; DB 6; Length 1041;
Best Local Similarity 60.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      5 RSHINPTGTV 14
Db      808 RQHLEPTGTI 817

RESULT 15
US-10-995-561-781
; Sequence 781, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 781
; LENGTH: 1097
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-781

Query Match          46.1%; Score 41; DB 6; Length 1097;
Best Local Similarity 60.0%; Pred. No. 19;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

Oy 5 RSHINPTGTV 14
| | : | | | :
Db 864 RQLEPTGTI 873

Search completed: January 1, 2006, 00:29:53
Job time : 1.91034 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:36:26 ; Search time 1.86207 Seconds
(without alignments)
799.198 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGVLLQL 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA*
1: /cgn2_6/ptodata/1/1aa/5 COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/PCTUS COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	18	2	US-09-626-127-9
2	89	100.0	18	2	US-09-993-059-29
3	89	100.0	18	2	US-10-103-327-29
4	89	100.0	22	2	US-09-626-127-7
5	89	100.0	22	2	US-09-993-059-27
6	89	100.0	22	2	US-10-103-327-27
7	89	100.0	24	2	US-09-626-127-10
8	89	100.0	24	2	US-09-993-059-30
9	89	100.0	24	2	US-10-103-327-30
10	89	100.0	26	2	US-09-626-127-5
11	89	100.0	26	2	US-09-993-059-25
12	89	100.0	26	2	US-10-103-327-25
13	89	100.0	28	2	US-09-626-127-8
14	89	100.0	30	2	US-09-626-127-3
15	89	100.0	30	2	US-09-993-059-23
16	89	100.0	30	2	US-10-103-327-23
17	89	100.0	32	2	US-09-626-127-6
18	89	100.0	32	2	US-09-993-059-26
19	89	100.0	32	2	US-10-103-327-26
20	89	100.0	36	2	US-09-626-127-4
21	89	100.0	36	2	US-09-993-059-24
22	89	100.0	36	2	US-10-103-327-24
23	89	100.0	92	1	US-07-602-824A-9
24	89	100.0	92	1	US-07-983-451-9
25	89	100.0	92	1	US-08-261-577-12
26	89	100.0	386	2	US-09-176-666-11
27	89	100.0	387	2	US-09-176-666-10

```

28      89 100.0 388 2 US-09-176-666-9 Sequence 9, Appli
29      89 100.0 389 2 US-09-176-666-8 Sequence 8, Appli
30      89 100.0 390 2 US-09-176-666-7 Sequence 7, Appli
31      89 100.0 391 2 US-09-176-666-6 Sequence 6, Appli
32      89 100.0 392 2 US-09-176-666-5 Sequence 5, Appli
33      89 100.0 393 2 US-09-176-666-4 Sequence 4, Appli
34      89 100.0 394 2 US-09-176-666-3 Sequence 3, Appli
35      89 100.0 396 2 US-09-176-666-2 Sequence 2, Appli
36      89 100.0 398 2 US-08-928-881-26 Sequence 26, Appli
37      89 100.0 398 2 US-09-176-666-1 Sequence 1, Appli
38      89 100.0 398 2 US-09-543-921-26 Sequence 26, Appli
39      89 100.0 398 2 US-09-266-014-4 Sequence 4, Appli
40      89 100.0 398 2 US-09-491-759-26 Sequence 26, Appli
41      89 100.0 398 2 US-10-360-101-202 Sequence 202, Appli
42      89 100.0 417 2 US-09-993-059-16 Sequence 16, Appli
43      89 100.0 417 2 US-10-103-327-16 Sequence 16, Appli
44      89 100.0 421 2 US-09-993-059-12 Sequence 12, Appli
45      89 100.0 421 2 US-10-103-327-12 Sequence 12, Appli

```

ALIGNMENTS

RESULT 1

```

US-09-626-127-9
; Sequence 9, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626.127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626.127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003.737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170.771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-9

```

```

Query Match      100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

```

RESULT 2

```

US-09-993-059-29
; Sequence 29, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993.059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0

```

```
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-29

Query Match      100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 3
US-10-103-327-29
; Sequence 29, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-29

Query Match      100.0%; Score 89; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 4
US-09-626-127-7
; Sequence 7, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1986-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-7

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 5
US-09-993-059-27
; Sequence 27, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-27

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 6
US-10-103-327-27
; Sequence 27, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-27

Query Match      100.0%; Score 89; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 7
US-09-626-127-10
```

; Sequence 10, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-10

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18

RESULT 8

US-09-993-059-30
; Sequence 30, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-30

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18

RESULT 9

US-10-103-327-30
; Sequence 30, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SEQ ID NO 30
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-30

Query Match 100.0%; Score 89; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 2.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18

RESULT 10

US-09-626-127-5
; Sequence 5, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-5

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGVLLQL 18
Db 1 TSLRSHINPTGVLLQL 18

RESULT 11

US-09-993-059-25
; Sequence 25, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25


```

; LENGTH: 26
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-25

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 12
US-10-103-327-25
; Sequence 25, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103.327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-25

Query Match 100.0%; Score 89; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 13
US-09-626-127-8
; Sequence 8, Application US/09626127
; Patent No. 6845968
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 28
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-8

Query Match 100.0%; Score 89; DB 2; Length 28;

```

```

Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 14
US-09-626-127-3
; Sequence 3, Application US/09626127
; Patent No. 6846968
; GENERAL INFORMATION:
; APPLICANT: Garget, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 0801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/09/626,127
; CURRENT FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-626-127-3

Query Match 100.0%; Score 89; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 15
US-09-993-059-23
; Sequence 23, Application US/099303059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGET, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 08010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-23

Query Match 100.0%; Score 89; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

```

```

Query Match      100.0%; Score 89; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  TSLRGRSHINPTGVLLQL 18
          |||||
Db      1  TSLRGRSHINPTGVLLQL 18

```

Search completed: December 31, 2005, 23:48:33
Job time : 1.86207 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:17 ; Search time 5.29655 Seconds
(without alignments)
1419.967 Million cell updates/sec

Title: US-10-602-220-29

Perfect score: 89

Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA_Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	18	3	US-09-993-059-29
2	89	100.0	18	4	US-10-103-327-29
3	89	100.0	18	4	US-10-602-219-29
4	89	100.0	18	4	US-10-602-220-29
5	89	100.0	18	4	US-10-684-300-9
6	89	100.0	18	4	US-10-684-349-9
7	89	100.0	18	5	US-10-851-388-29
8	89	100.0	18	5	US-10-984-389-29
9	89	100.0	22	3	US-09-993-059-27
10	89	100.0	22	4	US-10-103-327-27
11	89	100.0	22	4	US-10-602-219-27
12	89	100.0	22	4	US-10-602-220-27
13	89	100.0	22	4	US-10-684-300-7
14	89	100.0	22	4	US-10-684-349-7
15	89	100.0	22	5	US-10-851-388-27
16	89	100.0	22	5	US-10-984-389-27
17	89	100.0	24	3	US-09-993-059-30
18	89	100.0	24	4	US-10-103-327-30
19	89	100.0	24	4	US-10-602-219-30
20	89	100.0	24	4	US-10-602-220-30
21	89	100.0	24	4	US-10-684-300-10
22	89	100.0	24	4	US-10-684-349-10
23	89	100.0	24	5	US-10-851-388-30
24	89	100.0	24	5	US-10-984-389-30
25	89	100.0	26	3	US-09-993-059-25
26	89	100.0	26	4	US-10-103-327-25
27	89	100.0	26	4	US-10-602-219-25

28	89	100.0	26	4	US-10-602-220-25	Sequence 25, Appl
29	89	100.0	26	4	US-10-684-300-5	Sequence 5, Appl
30	89	100.0	26	4	US-10-684-349-5	Sequence 5, Appl
31	89	100.0	26	5	US-10-851-388-25	Sequence 25, Appl
32	89	100.0	26	5	US-10-984-389-25	Sequence 25, Appl
33	89	100.0	28	4	US-10-684-300-8	Sequence 8, Appl
34	89	100.0	28	4	US-10-684-349-8	Sequence 8, Appl
35	89	100.0	30	3	US-09-993-059-23	Sequence 23, Appl
36	89	100.0	30	4	US-10-103-327-23	Sequence 23, Appl
37	89	100.0	30	4	US-10-602-219-23	Sequence 23, Appl
38	89	100.0	30	4	US-10-602-220-23	Sequence 23, Appl
39	89	100.0	30	4	US-10-684-300-3	Sequence 3, Appl
40	89	100.0	30	4	US-10-684-349-3	Sequence 3, Appl
41	89	100.0	30	5	US-10-851-388-23	Sequence 23, Appl
42	89	100.0	30	5	US-10-984-389-23	Sequence 23, Appl
43	89	100.0	32	3	US-09-993-059-26	Sequence 26, Appl
44	89	100.0	32	4	US-10-103-327-26	Sequence 26, Appl
45	89	100.0	32	4	US-10-602-219-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-29
; Sequence 29, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-29

Query Match 100.0%; Score 89; DB 3; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
|||
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 2
US-10-103-327-29
; Sequence 29, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT

```
; ORGANISM: Tobacco mosaic virus
US-10-103-327-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

RESULT 3
US-10-602-219-29
; Sequence 29, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-219-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

RESULT 4
US-10-602-220-29
; Sequence 29, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
```

```
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR FILING DATE: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-220-29

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGVLLQL 18
Db 1 TSRLRSHINPTGVLLQL 18

RESULT 5
US-10-684-300-9
; Sequence 9, Application US/10684300
; Publication No. US20040064855A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,300
; CURRENT FILING DATE: 2003-10-09
; PRIOR FILING DATE: 09/626,127
; PRIOR APPLICATION NUMBER: 2000-07-26
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-300-9

Query Match      100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 6
US-10-684-349-9
; Sequence 9, Application US/10684349
; Publication No. US20040093646A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,349
; CURRENT FILING DATE: 2003-10-09
; PRIOR FILING DATE: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR FILING DATE: 1995-09-14
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-349-9

Query Match 100.0%; Score 89; DB 4; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 7
US-10-851-388-29
; Sequence 29, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-851-388-29

Query Match 100.0%; Score 89; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 8
US-10-984-389-29
; Sequence 29, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-984-389-29

Query Match 100.0%; Score 89; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 9
US-09-993-059-27
; Sequence 27, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-09-993-059-27

Query Match 100.0%; Score 89; DB 3; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
Db 1 TSRLRSHINPTGTVLLQL 18

RESULT 10
US-10-103-327-27
; Sequence 27, Application US/10103327
; Publication No. US20030108095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06

```
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-103-327-27
```

```
Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | |
Db 1 TSRLRSHINPTGVLLQL 18
```

```
RESULT 11
US-10-602-219-27
; Sequence 27, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-219-27
```

```
Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | |
Db 1 TSRLRSHINPTGVLLQL 18
```

```
RESULT 12
US-10-602-220-27
; Sequence 27, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-602-220-27
```

```
Query Match 100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0;
```

```
Qy 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | |
Db 1 TSRLRSHINPTGVLLQL 18
```

```
RESULT 13
US-10-684-300-7
; Sequence 7, Application US/10684300
; Publication No. US20040064855A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CFUS04
; CURRENT APPLICATION NUMBER: US/10/684,300
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
```



```
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-300-7

Query Match      100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 14
US-10-684-349-7
; Sequence 7, Application US/10684349
; Publication No. US20040093646A1
; GENERAL INFORMATION:
; APPLICANT: Garger, Stephen J.
; APPLICANT: Turpen, Thomas H.
; APPLICANT: KUMAGAI, Monto
; TITLE OF INVENTION: Production of lysosomal enzymes in plants by transient expression
; FILE REFERENCE: 00801.0087.CPUS04
; CURRENT APPLICATION NUMBER: US/10/684,349
; CURRENT FILING DATE: 2003-10-09
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/003,737
; PRIOR FILING DATE: 1995-09-14
; PRIOR APPLICATION NUMBER: 07/170,771
; PRIOR FILING DATE: 1988-03-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-684-349-7

Query Match      100.0%; Score 89; DB 4; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 15
US-10-851-388-27
; Sequence 27, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851.388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Tobacco mosaic virus
US-10-851-388-27
```

```
Query Match      100.0%; Score 89; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

Search completed: January 1, 2006, 00:29:25
Job time : 6.29655 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:15:26 ; Search time 6.24828 Seconds
(without alignments)
2032.482 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_eprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	89	100.0	403	2 Q6LER7 HUMAN	Q6LER7 homo sapien
2	89	100.0	429	1 AGAL HUMAN	P06280 homo sapien
3	89	100.0	429	2 Q53HF3 HUMAN	Q53HF3 homo sapien
4	89	100.0	429	2 Q53Y83 HUMAN	Q53Y83 homo sapien
5	52	58.4	809	2 Q92Z25 RHIME	Q92Z25 rhizobium m
6	50	56.2	419	1 AGAL_MOUSE	P51569 mus musculus
7	50	56.2	421	2 Q8BGZ6 MOUSE	Q8BGZ6 m mus muscu
8	47	52.8	206	2 Q5ND37 MOUSE	Q5ND37 mus musculus
9	47	52.8	312	2 Q4TAX9 TETNG	Q4TAX9 tetraodon n
10	47	52.8	417	1 PRDF_MOUSE	P97298 mus musculus
11	47	52.8	417	2 Q5ND38 MOUSE	Q5ND38 mus musculus
12	46	51.7	208	2 Q89KL6 BRAJA	Q89KL6 bradyrhizob
13	46	51.7	548	2 Q7SB68 NEUCR	Q7SB68 neurospora
14	46	51.7	609	2 Q22680 CAEEL	Q22680 caenorhabdi
15	45	50.6	213	2 Q7NXXD1 CHRVO	Q7NXXD1 chromobacte
16	45	50.6	219	2 Q5WXY9 LEGPL	Q5WXY9 legionella
17	45	50.6	219	2 Q5X615 LEGPA	Q5X615 legionella
18	45	50.6	219	2 Q5XZ14 LEGPH	Q5XZ14 legionella
19	45	50.6	221	2 Q82TY8 NITEU	Q82TY8 nitrosomona
20	45	50.6	240	2 Q8CMH3 STRA3	Q8CMH3 streptococc
21	45	50.6	306	2 Q7AMF9 NANEQ	Q7AMF9 nanocarchaeu
22	45	50.6	512	1 SPEE_OCEIH	Q8CV14 oceanobacil
23	45	50.6	542	1 RESEB_BACSU	P3161 bacillus su
24	44	49.4	184	2 Q8FPT8 COREF	Q8FPT8 corynebacte
25	44	49.4	224	2 Q6N4M2 RHOPA	Q6N4M2 rhodopsendo
26	44	49.4	542	2 Q62TC8 BACLD	Q62TC8 bacillus li
27	44	49.4	543	2 Q65HX9 BACLD	Q65HX9 bacillus li
28	44	49.4	723	2 Q82670 CICAR	Q82670 cicer ariet
29	43	48.3	183	2 Q8TCQ9_HUMAN	Q8TCQ9 homo sapien
30	43	48.3	196	1 Y1909_NEIMB	Q9JX82 neisseria m
31	43	48.3	286	1 AAC31_SALSP	P13246 salmonella

32	43	48.3	286	1 AAC32_SALSP	P0A255 salmonella
33	43	48.3	286	1 AAC32_ACIBA	P29807 acinetobact
34	43	48.3	286	1 AAC33_ENTCL	P0A256 enterobacte
35	43	48.3	286	2 Q52310_92ZZZ	Q52310 plasmid r.
36	43	48.3	286	2 Q5QJN0_SALTY	Q5QJN0 salmonella
37	43	48.3	286	2 Q7B8P8_PSEAE	Q7B8P8 pseudomonas
38	43	48.3	286	2 Q8GFQ5_CITFR	Q8GFQ5 citrobacter
39	43	48.3	286	2 Q79NT9_ACIBA	Q79NT9 acinetobact
40	43	48.3	286	2 Q03634_ECOLI	Q03634 escherichia
41	43	48.3	286	2 Q6LDA3_ECOLI	Q6LDA3 escherichia
42	43	48.3	296	1 Y608_TREPA	O83617 treponema p
43	43	48.3	338	2 Q9V2I0_PYRAB	Q9V2I0 pyrococcus
44	43	48.3	386	1 METK_FERTT	Q5NIC7 francisella
45	43	48.3	403	2 Q4TSL8_9SPHN	Q4TSL8 erythrobact

ALIGNMENTS

RESULT 1
Q6LER7 HUMAN PRELIMINARY; PRT; 403 AA.
AC Q6LER7;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Alpha-galactosidase A (BC 3.2.1.22) (Fragment).
GN Name=alpha-Gala;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Desnick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
DR EMBL; D00039; BAA34059.1; -; mRNA.
DR SMR; Q6LER7; 6-396.
DR GO; GO:0004557; F:alpha-galactosidase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDLASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycosidase; Hydrolase.
FT CHAIN 1 403 alpha-galactosidase A subunit peptide.
FT NON_TER 6 403
SQ SEQUENCE 403 AA; 45804 MW; 2FE193205BEB8D1A CRC64;

Query Match 100.0%; Score 89; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 1.7e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGTVLLQL 18
|||||
DB 374 TSRLRSHINPTGTVLLQL 391

RESULT 2
AGAL_HUMAN
ID_AGAL_HUMAN STANDARD; PRT; 429 AA.
AC P06280;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase
DE alfa).
GN Name=GLA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lymphoblast;
RX MEDLINE=89263745; PubMed=2542896;
RA Kornreich R., Desnick R.J., Bishop D.F.;
RT "Nucleotide sequence of the human alpha-galactosidase A gene.";
RL Nucleic Acids Res. 17:3301-3302(1989).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Fibroblast;
RX MEDLINE=87246603; PubMed=3036505;
RA Tsuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,
RA Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,
RA Ginns E.I.;
RT "Signal sequence and DNA-mediated expression of human lysosomal alpha-
RT galactosidase A".
RL Eur. J. Biochem. 165:275-280(1987).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci".
RL Mamm. Genome 6:334-338(1995).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15772651; DOI=10.1038/nature03440;
RA Ross M.T., Grafham D.V., Coffey A.J., Scherer S., McEay K., Muzny D.,
RA Platzer M., Howell G.R., Burrows C., Bird C.P., Frankish A.,
RA Lovell F.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,
RA Jones M.C., Hultes M.E., Andrews T.D., Scott C.E., Searle S.,
RA Ramser J., Whittaker A., Deadman K., Carter N.P., Hunt S.E., Chen R.,
RA Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,
RA Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,
RA Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhy S.,
RA Ashwell R.I., Babbage A.K., Baggeley C.L., Ballabio A., Banerjee R.,
RA Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,
RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
RA Chavez D., Chen E., Chen G., Chen Y., Chen Z., Chinault C.,
RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Clerc-Blankenburg K., Clifford K., Cobley J., Cole C.G., Conquer J.S.,
RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
RA Delgado O., Deshazo D., Dhani P., Ding Y., Dinh H., Dodsworth S.,
RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
RA Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L.,
RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
RA Griffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
RA Hawes A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
RA Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S.,
RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
RA Kioschis P., Klages S., Knights A.J., Kosiura A., Kovar-Smith C.,
RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
RA Lloyd C., Lloyd D.M., Loubege H., Loveland J.E., Lovell J.D.,
RA Lomad R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,

RA Okuwonu G., Palmer S., Pandian R., Parker D., Parrish J.,
RA Pasternak S., Patel D., Pearce A.V., Pearson D.M., Pelan S.E.,
RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
RA Ridler K.A., Schlessinger D., Schueler M.G., Sehra H.K.,
RA Shaw-Smith C., Shen H., Sheridan E.M., Shownkeen R., Skuce C.D.,
RA Smith M.L., Sotharan E.C., Steingruber H.E., Steward C.A., Storey R.,
RA Swann R.M., Thomas B., Thoburn D., Tabor P.E., Taudien S., Taylor T.,
RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
RA Tromans A.C., d'Urso M., Verduzco D., Villasana D., Waldron L.,
RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
RA Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L.,
RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
RA Zorilla S., Buck D., Reinhardt R., Poustka A., Rosenthal A.,
RA Lehrach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
RA Nelson D.B., Weinstein G., Sulston J.E., Durbin R., Hubbard T.,
RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
RT "The DNA sequence of the human X chromosome.";
RL Nature 434:325-337(2005).
RN [5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Myers R.M.,
RA Butterfield Y.S.N., Kryzyski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Desnick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863(1986).
RN [7]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=88112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
RT human lysosomal enzyme, alpha-galactosidase A.";
RL Gene 58:177-188(1987).
RN [8]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=88234528; PubMed=2836863;
RA Bishop D.F., Kornreich R., Desnick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
RT further evidence for the absence of a 3' untranslated region.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907(1988).
RN [9]
RP RNA EDITING OF POSITION 396.
RX MEDLINE=95380278; PubMed=7503918;
RA Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
RT purine conversion.";

RL Nucleic Acids Res. 23:2636-2640(1995).
RN [10]
RP X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
RC PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
RX PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
RA Garman S.C., Garboczi D.N.;
RT "The molecular defect leading to Fabry disease: structure of human
alpha-galactosidase.";
RL J. Mol. Biol. 337:319-335(2004).
RN [11]
RP REVIEW ON FD VARIANTS.
RX MEDLINE=94258158; PubMed=7911050;
RA Eng C.M., Desnick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
human alpha-galactosidase A gene.";
RN Hum. Mutat. 3:103-111(1994).
RN [12]
RP VARIANT FD SER-40.
RX MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
RT activity caused by a single amino acid substitution of Pro-40 by
Ser.";
RL FEBS Lett. 259:353-356(1990).
RN [13]
RP VARIANT FD VAL-296.
RX MEDLINE=91101674; PubMed=1846223;
RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Desnick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
to the myocardium.";
RL N. Engl. J. Med. 324:395-399(1991).
RN [14]
RP VARIANT FD GLN-301.
RX MEDLINE=91022721; PubMed=2171331;
RA Sakuraba H., Oshima A., Fukuhara Y., Shimmoto M., Nagao Y.,
Bishop D.F., Desnick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
in classical and atypical hemizygotes with Fabry disease.";
RL Am. J. Hum. Genet. 47:784-789(1990).
RN [15]
RP VARIANT FD TRP-356.
RX MEDLINE=89198098; PubMed=2539398;
RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
Sakuraba H., Desnick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
in the alpha-galactosidase gene.";
RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]
Query Match 100.0%; Score 89; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 3
Q53HF3 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
Query Match 100.0%; Score 89; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 3
Q53HF3 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53HF3;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;

RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
RT "Oligo-capping: a simple method to replace the cap structure of
RT eucaryotic mRNAs with oligoribonucleotides.";
RL Gene 138:171-174(1994).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
RT end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK222627; BAD96347.1; -; mRNA.
FT NON TER 1
SQ SEQUENCE 429 AA; 48752 MW; AFA828F5B22ED76 CRC64;
Query Match 100.0%; Score 89; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 4
Q53Y83 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53Y83;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Kainine N., Chen X., Rolfs A., Halleck A., Hines L., Eisenstein S.,
Koundinya M., Raphael J., Moreira D., Kelley T., LaBaer J., Lin Y.,
Phelan M., Farmer A.;
RT "Cloning of human full-length cDNAs in BD Creator(TM) System Donor
RT vector.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BT006864; AAP35510.1; -; mRNA.
SQ SEQUENCE 429 AA; 48766 MW; 613F8BF21B107D7B CRC64;
Query Match 100.0%; Score 89; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 TSRLRSHINPTGTVLLQL 18
Db 400 TSRLRSHINPTGTVLLQL 417
RESULT 5
Q92Z25 RHIME PRELIMINARY; PRT; 809 AA.
AC Q92Z25;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

```

DE Adenylate cyclase, putative.
GN OrderedLocNames=RA0303; ORFNames=SMa0579;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432; DOI=10.1073/pnas.161294798;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huisar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AF007222; XAK64961.1; -; Genomic_DNA.
DR FPIR; G95299;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0000287; F:magnesium ion binding; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006171; F:cAMP biosynthesis; IEA.
DR GO; GO:0007242; F:intracellular signaling cascade; IEA.
DR InterPro; IPR01054; G:Cyclase.
DR InterPro; IPR003660; His_kin_HAMP.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR Pfam; PF00672; HAMP; 1.
DR SMART; SM00044; CYCC; 1.
DR SMART; SM00304; HAMP; 1.
DR PROSITE; PS0125; GUANYLATE_CYCLASES_2; 1.
DR PROSITE; PS00885; HAMP; 1.
KW Complete proteome; Plasmid.
SQ SEQUENCE 809 AA; 88948 MW; F650FFE4BD67BA21 CRC64;

Query Match 58.4%; Score 52; DB 2; Length 809;
Best Local Similarity 66.7%; Pred. No. 9.2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 RLRSHPNTGTVLQ 17
||:|||||:|
Db 486 RLEAHRNPTGTLQ 500

RESULT 6
AGAL_MOUSE STANDARD; PRT; 419 AA.
AC P51569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A).
GN Name=Gla; Synonyms=Ags;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Kidney;
RX MEDLINE=96125203; PubMed=85431175; DOI=10.1016/0378-1119(95)00592-7;
RA Ohshima T., Murray G.J., Negie J.W., Quirk J.M., Kraus M.H.,
RA Barton N.W., Brady R.O., Kulkarni A.B.;
RT "Structural organization and expression of the mouse gene encoding
RT alpha-galactosidase A."
RL Gene 166:277-280(1995).

```

```

RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C129;
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.-C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci."
RL Mamm. Genome 6:334-338(1995).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96316016; PubMed=8733892; DOI=10.1006/bmme.1996.0020;
RA Gotlib R.W., Bishop D.F., Wang A.M., Zeidner K.M., Ioannou Y.I.,
RA Adler D.A., Distèche C.M., Deanick R.J.;
RT "The entire genomic sequence and cDNA expression of mouse alpha-
RT galactosidase A."
RL Biochem. Mol. Med. 57:139-148(1996).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing alpha-D-
CC galactose residues in alpha-D-galactosides, including galactose
CC oligosaccharides, galactomannans and galactohydrolase.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
-----
This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
-----
EMBL; U34071; AAA96749.1; -; mRNA.
EMBL; L46651; AAA74453.1; -; Genomic DNA.
EMBL; U58105; ABA47244.1; -; Genomic DNA.
EMBL; U50716; AAC52584.1; -; mRNA.
EMBL; U50715; AAC52583.1; -; Genomic DNA.
EMBL; BC009021; AAH09021.1; -; mRNA.
PIR; JC4522; JC4522.
SMR; P51569; 32-418.
Ensembl; ENSMUSG000000031266; Mus musculus.
MGI; MGI:1347344; Gla.
GO; GO:0005615; C:extracellular space; TAS.
InterPro; IPR002241; Glyco_hydro_27.
InterPro; IPR000111; Glyco_hydro_GHD.
Pfam; PF02065; Melibiase; 1.
PRINTS; PR00740; GLHYDRLASE27.
PRODOM; PD005572; Glyco_hydro_GHD; 1.
PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycoprotein; Glycosidase; Hydrolase; Lysosome; Signal.
FT SIGNAL 1 31

```

```

FT CHAIN 32 419 Alpha-galactosidase A.
FT ACT SITE 170 Nucleophile (By similarity).
FT ACT SITE 231 Proton donor (By similarity).
FT SITE 231 Substrate binding (By similarity).
FT CARBOHYD 139 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 192 N-linked (GlcNAc...) (By similarity).
FT CARBOHYD 215 N-linked (GlcNAc...) (By similarity).
FT DISULFID 52 By similarity.
FT DISULFID 56 By similarity.
FT DISULFID 142 By similarity.
FT DISULFID 202 By similarity.
FT DISULFID 378 By similarity.
SQ SEQUENCE 419 AA; 47643 MW; BDSE6A99AC113613 CRC64;

Query Match 56.2%; Score 50; DB 1; Length 419;
Best Local Similarity 50.0%; Pred. No. 9.5;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSLRSHNPQTGLVQL 18
DB 400 TLTLRVNSPTGLFRL 417

RESULT 7
QBEGZ6 MOUSE
ID QBEGZ6 MOUSE PRELIMINARY; PRT; 421 AA.
AC QBEGZ6;
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430057F16 product:galactosidase, alpha, full insert
DE sequence (Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
DE full-length enriched library, clone:E330039P08 product:galactosidase,
DE alpha, full insert sequence).
CN Name=Gla;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Baldarelli R., Barsh G.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleisemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Mateuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staab F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bernaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guastincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Marzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).

```

```

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,
RA Nikaide I., Oeato N., Saito R., Suzuki H., Yananaka I., Gojobori T.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Guastincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai J., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pettea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C.F., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Ovary, and Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hasegawa T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu M., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Oheato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashi S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK040081; BAC30508.1; -; mRNA.
DR EMBL; AK054547; BAC35819.1; -; mRNA.
DR HSSP; P06280; 1R46.

```

```

DR SMR: Q8BG26; 34-420.
DR MGI:1347344; Glia.
DR GO:GO:0005615; C:extracellular space; TAS.
DR InterPro: IPR002241; Glyco_hydro_27.
DR Pfam: PF02065; Melibiase; 1.
DR PRINTS: PR00740; GLHYDRLASE27.
DR PRODOM: PD002572; Glyco_hydro_GHD; 1.
DR PROSITE: PS00512; ALPHA GALACTOSIDASE; 1.
DR SEQUENCE 421 AA; 47844 MW; 5F66772334014B6F CRC64;

Query Match 56.2%; Score 50; DB 2; Length 421;
Best Local Similarity 50.0%; Pred. No. 9.6;
Matches 9; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSLRLSHINPTGVLLQL 18
| : : : : :
Db 402 TLLTKTRVPSGTVLFLR 419

RESULT 8
Q5ND37 MOUSE PRELIMINARY; PRT; 206 AA.
AC Q5ND37;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Serine (Or cysteine) proteinase inhibitor, clade F, member 1
DE (Fragment).
GN Name=Serpinfl; ORFNames=RP23-384C18.1-002;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RA Phillimore B.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL591496; CAI35379.1; -; Genomic DNA.
DR GO:GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
DR InterPro: IPR00215; Prot_inh_serpin.
DR Pfam: PF00079; Serpin; 2.
DR NON_TER 206 206
FT SEQUENCE 206 AA; 22551 MW; F24295B0571CCC5C CRC64;

Query Match 52.8%; Score 47; DB 2; Length 206;
Best Local Similarity 71.4%; Pred. No. 14;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 RLRSHINPTGVLL 16
| : : : : :
Db 66 RLRSASPTGNVLL 79

RESULT 9
Q4TAX9 TETNG PRELIMINARY; PRT; 312 AA.
AC Q4TAX9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF7240, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00004000001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthopterygia; Acanthopterygii; Percormpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]_
RP NUCLEOTIDE SEQUENCE.

```

```

RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani V., Katinka M., Vacherie B.,
RA Biewont C., Skalli Z., Cattolico L., Poullain J., De Bernardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutancieu J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -!- SIMILARITY: Belongs to the TGF-beta family.
DR EMBL; CAAE01007240; CAF89953.1; -; Genomic DNA.
DR InterPro: IPR001839; TGFb.
DR Pfam: PF00019; TGF_beta; 1.
DR PRODOM: PD000357; TGFb; 1.
DR SMART; SM00204; TGFb; 1.
DR PROSITE; PS00250; TGF_BETA_1; 1.
KW Growth factor.
FT NON_TER 1 1
FT NON_TER 312 312
FT SEQUENCE 312 AA; 33461 MW; D9AABD81B0B9FD79 CRC64;

Query Match 52.8%; Score 47; DB 2; Length 312;
Best Local Similarity 53.3%; Pred. No. 22;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 RLRSHINPTGVLLQ 17
| : : : : :
Db 252 RLRLSHLEPTNHAIQ 266

RESULT 10
PEDF_MOUSE STANDARD; PRT; 417 AA.
AC P97298; O70629; O88691;
DT 15-JUL-1998 (Rel. 36, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Pigment epithelium-derived factor precursor (PEDF) (Stromal cell-
DE derived factor 3) (SDF-3) (Caspin).
GN Name=Serpinfl; Synonym=Pedf; Sdf3;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=97092876; PubMed=8938438; DOI=10.1006/geno.1996.0560;
RA Shirozu M., Tada H., Tashiro K., Nakamura T., Lopez N.D., Nazarea M.,
RA Hamada T., Sato T., Nakano T., Honjo T.;
RT "Characterization of novel secreted and membrane proteins isolated by
RT the signal sequence trap method.";
RL Genomics 37:273-280(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE. PROTEIN SEQUENCE OF 53-66; 252-261; 333-344 AND
RP 359-372. TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX STRAIN=BAALB/c; TISSUE=Liver;
RX MEDLINE=98279032; PubMed=9614124; DOI=10.1074/jbc.273.24.15125;
RA Kozaki K., Miyaishi O., Koizumi Y., Yasui Y., Kashiwa A.,
RA Nishikawa Y., Shimizu S., Saga S.;

```


RT "Isolation, purification and characterization of a collagen-associated
 RT serpin, capsin, produced by murine colon adenocarcinoma cells";
 J. Biol. Chem. 273:15125-15130(1998).
 [3]
 RN NUCLEOTIDE SEQUENCE, AND TISSUE SPECIFICITY.
 RP STRAIN=FVB/N; TISSUE=Liver;
 RC PubMed=59565647;
 RA Singh V.K., Chader G.J., Rodriguez I.R.;
 RX "Structural and comparative analysis of the mouse gene for pigment
 RT epithelium-derived factor (PEDF).";
 RL Mol. Vis. 4:7-7(1998).
 [4]
 RN NUCLEOTIDE SEQUENCE.
 RP STRAIN=BA1B/c;
 RC Tombran-Tink J.;
 RA Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 [5]
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RP STRAIN=FVB/N; TISSUE=Liver;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Viallon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.L., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [6]
 RN FUNCTION.
 RP PubMed=12632345; DOI=10.1053/jpsu.2003.50104;
 RX Abramson L.P., Stellmach V., Doll J.A., Cornwell M., Arensman R.M.,
 RA Crawford S.E.;
 RT "Wilms' tumor growth is suppressed by antiangiogenic pigment
 RP epithelium-derived factor in a xenograft model.";
 RL J. Pediatr. Surg. 38:336-342(2003).
 CC -!- FUNCTION: Neurotrophic protein; induces extensive neuronal
 CC differentiation in retinoblastoma cells. Potent inhibitor of
 CC angiogenesis. As it does not undergo the S (stressed) to R
 CC (relaxed) conformational transition characteristic of active
 CC serpins, it exhibits no serine protease inhibitory activity.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Highly expressed in the liver, gastric
 CC glandular mucosa and renal tubules. It is also expressed in the
 CC brain, heart, lung retina and testes.
 CC -!- DEVELOPMENTAL STAGE: First detected at 12.5 dpc in cartilage
 CC primordium, it is present in the osseous matrix of developing
 CC limbs, vertebrae, ribs and skull. At 16.5 dpc it is detected in
 CC bone matrix and smooth muscle, and at lower levels in connective
 CC tissue, bronchial epithelial cells, metanephron microtubules, and
 CC skin.
 CC -!- SIMILARITY: Belongs to the serpin family.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 CC
 CC EMBL; D50460; BAA09051.1; -; mRNA.
 CC EMBL; AF036164; AAC69271.1; -; mRNA.

DR EMBL; D87975; BAA31978.1; -; mRNA.
 DR EMBL; AF017055; AAC05733.1; -; Genomic DNA.
 DR EMBL; AF017051; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017052; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017053; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017054; AAC05733.1; JOINED; Genomic DNA.
 DR EMBL; AF017057; AAC05731.1; -; mRNA.
 DR EMBL; BC019852; AAH19852.1; -; mRNA.
 DR HSSP; P36955; 1IMV.
 DR Ensemble; ENSMUSG0000000753; Mus musculus.
 DR MGI; MGI:108080; Serpinf1.
 DR GO; GO:0005615; C:extracellular space; IDA.
 DR GO; GO:0016525; P:negative regulation of angiogenesis; IDA.
 DR GO; GO:0050769; P:positive regulation of neurogenesis; ISS.
 DR InterPro; IPR000215; Prot_inh_serpin.
 DR PANTHER; PTHR11461; Prot_inh_serpin; 1.
 DR Pfam; PF00079; Serpin; 1.
 DR SMART; SM00093; SERPIN; 1.
 DR PROSITE; PS00284; SERPIN; 1.
 KW Direct protein sequencing; Glycoprotein; Serpin; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 417
 FT CARBOHYD 284 284
 FT N-linked (GlcNAc...) (Potential).
 FT CONFLICT 70 70
 FT S -> G (in Ref. 4).
 FT CONFLICT 136 136
 FT K -> N (in Ref. 2).
 FT CONFLICT 137 137
 FT S -> R (in Ref. 4).
 FT CONFLICT 280 280
 FT T -> A (in Ref. 1).
 FT CONFLICT 377 377
 FT Q -> L (in Ref. 4).
 SQ SEQUENCE 417 AA; 46234 MW; ECD360FEGAA74D25 CRC64;
 Query Match 52.8%; Score 47; DB 1; Length 417;
 Best Local Similarity 71.4%; Pred. No. 31;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 3 RLRSHINPTGVLL 16
 Db 66 RLRSASPTGNVLL 79
 RESULT 11
 QSN38 MOUSE
 ID QSN38_MOUSE PRELIMINARY; PRT; 417 AA.
 AC QSN38;
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Serine (Or cysteine) proteinase inhibitor, clade F, member 1.
 GN Name=Serpinf1; ORFNames=RP23-384C18.1-001;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Phillimore B.;
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to the serpin family.
 DR EMBL; AL591496; CA135378.1; -; Genomic DNA.
 DR GO; GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
 KW Serpin.
 SQ SEQUENCE 417 AA; 46234 MW; ECD360FEGAA74D25 CRC64;
 Query Match 52.8%; Score 47; DB 2; Length 417;
 Best Local Similarity 71.4%; Pred. No. 31;
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 Qy 3 RLRSHINPTGVLL 16
 Db 66 RLRSASPTGNVLL 79
 RESULT 12

```

Q89KL6 BRAJA
ID Q89KL6_BRAJA PRELIMINARY; PRT; 208 AA.
AC Q89KL6;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Transcriptional regulatory protein.
GN OrderedLocusNames=blr4889;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idegawa K., Iriiguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197(2002).
CC -!- SIMILARITY: Contains 1 HTH LysR-type DNA-binding domain.
DR EMBL; BA000040; BAC50154.1; -; Genomic_DNA.
DR HSPF; Q9WCX7; IIXC.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR000847; HTH LysR.
DR InterPro; IPR011991; Wing_hlx_DNA_bd.
DR Pfam; PF0126; HTH_1; 1.
DR PRINTS; PR00039; HTHLYSR.
DR PROSITE; PS0931; HTH LYSR; 1.
KW Complete proteome; DNA-binding; Transcription;
KW Transcription regulation.
SQ SEQUENCE 208 AA; 22591 MW; 5C255F49DAC0D5DB CRC64;

Query Match 51.7%; Score 46; DB 2; Length 208;
Best Local Similarity 69.2%; Pred. No. 21;
Matches 9; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 SHINPTGVLLQL 18
| : : : : :
DB 85 SOLKPSGVLLQL 97

RESULT 13
Q7SB68 NEUCR
ID Q7SB68_NEUCR PRELIMINARY; PRT; 548 AA.
AC Q7SB68;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU06273.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenkoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysellis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,

```

```

RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nubaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC preliminary data.
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000173; EAA33641.1; -; Genomic_DNA.
SQ SEQUENCE 548 AA; 61216 MW; 7E23FDFD00D83219 CRC64;

Query Match 51.7%; Score 46; DB 2; Length 548;
Best Local Similarity 52.9%; Pred. No. 63;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQ 17
| : : : : :
DB 399 TAALRRYLNPPTTVRL 415

RESULT 14
Q22680 CAEEL
ID Q22680 CAEEL PRELIMINARY; PRT; 609 AA.
AC Q22680;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Hypothetical protein T22C8.7.
GN ORFNames=T22C8.7;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=93069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z49071; CAA88879.1; -; Genomic_DNA.
DR PIR; T25120; T25120.
DR Ensembl; T22C8.7; Caenorhabditis elegans.
DR WormBase; WBGene0011928; T22C8.7.
DR WormPep; T22C8.7; CE02355.
DR InterPro; IPR001507; Endoglin/CD105.
DR Pfam; PF00100; Zona_pellucida; 1.
DR SMART; SM00241; ZP_1.
DR PROSITE; PS51034; ZP_2; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 609 AA; 68883 MW; 1DC9BA97CAFBAC2D CRC64;

Query Match 51.7%; Score 46; DB 2; Length 609;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 LRSHINPTGT 13
| : : : : :
DB 43 IRLHINPTGT 52

RESULT 15
Q7NXXD1 CHRVO
ID Q7NXXD1_CHRVO PRELIMINARY; PRT; 213 AA.
AC Q7NXXD1;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Probable integral membrane protein.
GN OrderedLocusNames=CV1696;
OS Chromobacterium violaceum.

```

OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Chromobacterium.
 OX NCBI_TaxID=536;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ATCC 12472 / DSM 30191;
 RX MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
 RA Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
 RA Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
 RA Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
 RA Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
 RA Batista J.S., Belo A., van den Berg C., Bogo M., Bonatto S.,
 RA Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
 RA Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carriaro D.M.,
 RA Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
 RA Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
 RA Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
 RA Ferro M.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
 RA Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
 RA Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
 RA Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyza M.C.C.P.,
 RA Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
 RA di Mauro S.M.Z., de Medeiros S.R.B., Meisner R.V., Moreira M.A.M.,
 RA Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
 RA Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
 RA Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
 RA Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
 RA Santos E.B.P., Santos F.R., Schneider M.P.C., Seunanez H.N.,
 RA Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
 RA Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
 RA Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
 RA Vettore A., Wassen R., Zaha A., Simpson A.J.G.;
 RT "The complete genome sequence of Chromobacterium violaceum reveals
 RT remarkable and exploitable bacterial adaptability";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
 DR EMBL; AE016916; AAQ59371.1; -, Genomic DNA.
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR008915; Peptidase M50.
 DR Pfam; PF02163; Peptidase_M50; 1.
 KW Complete proteome.
 SQ SEQUENCE 213 AA; 23777 MW; 1A941A4603F0AC1F CRC64;

Query Match 50.6%; Score 45; DB 2; Length 213;
 Best Local Similarity 75.0%; Pred. No. 32;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 7 HINPTGVLLQL 18
 ||| |||||
 Db 54 HIDPIGTVLLPL 65

Search completed: December 31, 2005, 23:47:04
 Job time : 8.24828 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:16:51 ; Search time 1.44828 Seconds
(without alignments)
1195.837 Million cell updates/sec

Title: US-10-602-220-29
Perfect score: 89
Sequence: 1 TSRLRSHINPTGTVLLQL 18
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	100.0	429	1 GBHUA	alpha-galactosidase
2	52	58.4	809	2 G95299	probable adenylate
3	50	56.2	419	2 JG4522	alpha-galactosidase
4	46	51.7	609	2 T35120	hypothetical prote
5	45	50.6	542	2 S45557	resb protein - bac
6	43	48.3	196	2 D81028	Mat/YceF/YhdE fami
7	43	48.3	286	2 J80138	gentamicin-N-acety
8	43	48.3	286	2 S06032	gentamycin 3'-N-ac
9	43	48.3	286	2 S06030	aminoglycoside N3'
10	43	48.3	286	2 S09651	gentamicin resist
11	43	48.3	286	2 S14541	hypothetical prote
12	43	48.3	286	2 JG2063	hypothetical prote
13	43	48.3	296	2 B71305	bone morphogenetic
14	43	48.3	338	2 C75196	bone morphogenetic
15	42	47.2	125	2 S43295	bone morphogenetic
16	42	47.2	151	2 S43296	bone morphogenetic
17	42	47.2	243	2 A33329	testis-specific pr
18	42	47.2	289	2 T34241	hypothetical prote
19	42	47.2	436	2 B55452	cartilage-derived
20	42	47.2	556	2 B83847	cytochrome c bioge
21	41	46.1	85	2 S04808	hypothetical prote
22	41	46.1	114	2 T10111	hypothetical prote
23	41	46.1	123	1 BKIP	biotin carboxyl ca
24	41	46.1	267	1 DCVKOP	biotidine-5'-phos
25	41	46.1	347	2 A90318	hypothetical prote
26	41	46.1	361	2 S57895	hypothetical prote
27	41	46.1	495	2 S43294	bone morphogenetic
28	41	46.1	501	2 JG2347	growth/differentia
29	41	46.1	501	2 A55452	cartilage-derived

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N/Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1987 #sequence_revision 27-Oct-1995 #text_change 09-Jul-2004
C/Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R/Kornreich, R.; Desnick, R.J.; Bishop, D.F.

Nucleic Acids Res. 17, 3301-3302, 1989
A/Title: Nucleotide sequence of the human alpha-galactosidase A gene.

A/Reference number: S04081; MUID:89263745; PMID:2542896

A/Accession: S04081

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-429 <KOR>

A/Cross-references: UNIPROT:P06280; UNIPARC:UPI00000033A30; EMBL:X14448; NID:g11755; PID:

R/Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.

Gene 58, 177-188, 1987

A/Title: A genomic clone containing the promoter for the gene encoding the human lysosom

A/Reference number: A29608; MUID:88112869; PMID:2892762

A/Accession: A29608

A/Molecule type: DNA

A/Residues: 1-64 <QOI>

A/Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:

R/Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A/Title: Structural organization of the human alpha-galactosidase A gene: further eviden

A/Reference number: A30214; MUID:88234528; PMID:2836863

A/Accession: A30214

A/Molecule type: DNA

A/Residues: 1-64 <BIS>

A/Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R/Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A/Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity c

A/Reference number: S14879; MUID:90092580; PMID:2152885

A/Accession: S14879

A/Molecule type: mRNA

A/Residues: 1, 'K', 'J', '3', '39', 'S', '41', '429' <KOI>

A/Cross-references: UNIPARC:UPI000011E285; EMBL:X16889

A/Experimental source: Fabry's disease patient

R/Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick, R

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A/Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the r

A/Reference number: A00896; MUID:86259694; PMID:3014515

A/Accession: A00896

A/Molecule type: mRNA

A/Residues: 27-429 <B12>

A/Cross-references: UNIPARC:UPI0000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID

A/Experimental source: lung

A/Accession: B00896

A/Molecule type: protein

A;Residues: 1-542 <SOR>
A;Cross-references: UNIPROT:P35161; UNIPARC:UPI0000060679; EMBL:L09228; NID:g410114; PID R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berteaux C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E. Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallez lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hulio, M.F. Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Ross, M.; Sadaie, Y.; Sato, T.; Scanlon, A;Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron akuechi, M.; Tamakoshi, A.; Tanaka, T.; Terptratr, A.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69691
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-542 <KUN>
A;Cross-references: UNIPARC:UPI0000060679; GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CA A;Experimental source: strain 168
C;Genetics:
A;Gene: resB

Query Match 50.6%; Score 45; DB 2; Length 542;
Best Local Similarity 72.7%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 7 HINPTGTLLQ 17
|||||
Db 10 HINPVGTVLCE 20
|||||

RESULT 6
D81028
Maf/Ycef/Yhde family protein NMB1909 [imported] - Neisseria meningitidis (strain MC58 se C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
C;Accession: D81028
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; xi, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: D81028
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-196 <TEF>
A;Cross-references: UNIPROT:Q9JXS2; UNIPARC:UPI000013B498; GB:AE002540; GB:AE002098; NID A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1909
C;Superfamily: septum formation protein maf

Query Match 48.3%; Score 43; DB 2; Length 196;
Best Local Similarity 50.0%; Pred. NO. 11;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 TSLRSHINPTGTLLQL 18
|||||
Db 116 TGRMRRHIDKTVVNRQL 133
|||||

RESULT 7
JE0138
gentamicin-N-acetyltransferase(3)-V - human
C;Species: Homo sapiens (man)
C;Date: 10-Jul-1998 #sequence_revision 10-Jul-1998 #text_change 26-May-2000

C;Accession: JE0138
R;Fang, Y.; Song, H.Y.; Wang, F.; Zhu, D.Z.
Chinese J. Microbiol. Immunol. 16, 409-412, 1996
A;Title: Molecular cloning and characterization of gentamicin-N-acetyltransferase (3)-V
A;Reference number: JE0138
A;Accession: JE0138
A;Molecule type: mRNA
A;Residues: 1-286 <FAN>
A;Cross-references: UNIPARC:UPI00001794CE
C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
|||||
Db 45 LRSVAGPTGTVM 56
|||||

RESULT 8
S06032
gentamycin 3'-N-acetyltransferase (EC 2.3.1.60) type IV - Enterobacteriaceae plasmid pWE C;Species: plasmid pWP14a
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
C;Accession: S06032
R;Allmangberger, R.; Braeu, B.; Piepersberg, W.
Mol. Gen. Genet. 198, 514-520, 1985
A;Title: Genes for gentamicin-(3)-N-acetyl-transferases III and IV. II. Nucleotide sequence A;Reference number: S06030; MUID:85239912; PMID:3892230
A;Accession: S06032
A;Molecule type: DNA
A;Residues: 1-286 <ALL>
A;Cross-references: UNIPARC:UPI0000000849; EMBL:X13542; NID:g45940; PIDN:CAA31893.1; PIC C;Genetics:
A;Gene: aacC3
A;Genome: plasmid pWP14a
C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C;Keywords: acyltransferase; antibiotic resistance; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
|||||
Db 45 LRSVAGPTGTVM 56
|||||

RESULT 9
S06030
gentamycin 3'-N-acetyltransferase (EC 2.3.1.60) type III - Enterobacteriaceae plasmid pW C;Species: Enterobacteriaceae
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
C;Accession: S06030
R;Allmangberger, R.; Braeu, B.; Piepersberg, W.
Mol. Gen. Genet. 198, 514-520, 1985
A;Title: Genes for gentamicin-(3)-N-acetyl-transferases III and IV. II. Nucleotide sequence A;Reference number: S06030; MUID:85239912; PMID:3892230
A;Accession: S06030
A;Molecule type: DNA
A;Residues: 1-286 <ALL>
A;Cross-references: UNIPARC:UPI000012508C; EMBL:X13543; NID:g45943; PIDN:CAA31895.1; PIC C;Genetics:
A;Gene: aacC3
A;Genome: plasmid pWP13a
C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C;Keywords: acyltransferase; antibiotic resistance; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
||| : |||||
Db 45 LRSVAGPTGTVM 56

RESULT 10
S09651
aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) isozyme II - Enterobacter cloacae
C;Species: Enterobacter cloacae
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: S09651
R;Vliegenthart, J.S.; Ketelaar-van Gaalen, P.A.G.; van de Klundert, J.A.M.
Antimicrob. Agents Chemother. 33, 1153-1159, 1999
A;Title: Nucleotide sequence of the aacC2 gene, a gentamicin resistance determinant invc
A;Accession: S09651
A;Molecule type: DNA
A;Residues: 1-286 <VLI>
A;Cross-references: UNIPROT:P13245; UNIPARC:UPI0000000849; EMBL:X51534; NID:G40878; PIDN
C;Genetics:
A;Gene: aacC2
C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C;Keywords: acyltransferase; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
||| : |||||
Db 45 LRSVAGPTGTVM 56

RESULT 11
S14541
aminoglycoside N3'-acetyltransferase (EC 2.3.1.81) - Escherichia coli
C;Species: Escherichia coli
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S14541
R;Vakulenko, S.B.; Entina, H.
submitted to the EMBL Data Library, October 1990
A;Description: Nucleotide sequence of the aacC2 gene from E. coli R-plasmid.
A;Reference number: S14541
A;Accession: S14541
A;Molecule type: DNA
A;Residues: 1-286 <VAK>
A;Cross-references: UNIPROT:Q03634; UNIPARC:UPI000000B81C6; EMBL:X54723; NID:G45769; PIDN
C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase
C;Keywords: acyltransferase; coenzyme A

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
||| : |||||
Db 45 LRSVAGPTGTVM 56

RESULT 12
JC2063
gentamicin resistance determinant - Serratia marcescens
C;Species: Serratia marcescens
C;Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 09-Jul-2004
C;Accession: JC2063
R;Jung, J.S.; Cheong, T.C.; Cho, M.S.; Hah, Y.C.; Chung, J.H.
Biochem. Biophys. Res. Commun. 198, 1084-1089, 1994
A;Title: Nucleotide sequence and expression of a gentamicin resistance gene isolated fro
A;Reference number: JC2063; MUID:94161718; PMID:8117265
A;Accession: JC2063
A;Molecule type: DNA
A;Residues: 1-286 <JUN>
A;Cross-references: UNIPROT:Q52310; UNIPARC:UPI000000B39F5

C;Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

Query Match 48.3%; Score 43; DB 2; Length 286;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 LRSHINPTGTVL 15
||| : |||||
Db 45 LRSVAGPTGTVM 56

RESULT 13

B71305

hypothetical protein TP0608 - syphilis spirochete

C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004

C;Accession: B71305

R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin

rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDo

they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.

Science 281, 375-388, 1998

A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A;Reference number: A71250; MUID:98332770; PMID:9665876

A;Accession: B71305

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-296 <COL>

A;Cross-references: UNIPROT:O83617; UNIPARC:UPI00000139B22; GB:AE001235; GB:AE000520; NID

A;Experimental source: strain Nichols

C;Genetics:

A;Gene: TP0608

C;Superfamily: syphilis spirochete hypothetical protein TP0608

Query Match 48.3%; Score 43; DB 2; Length 296;
Best Local Similarity 52.9%; Pred. No. 17;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SRLRSHINPTGTVLLQL 18
||| : ||||| : |||
Db 230 SRKFSAINPTGSLALEI 246

RESULT 14

C75196

hypothetical protein PAB0054 - Pyrococcus abyssi (strain Orsay)

C;Species: Pyrococcus abyssi

C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C;Accession: C75196

R;anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A;Reference number: A75001

A;Accession: C75196

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-338 <KAW>

A;Cross-references: UNIPROT:Q9V2I0; UNIPARC:UPI00000034415; GB:AJ248283; GB:AL096836; NID

A;Experimental source: strain Orsay

C;Genetics:

A;Gene: PAB0054

C;Superfamily: Pyrococcus abyssi hypothetical protein PAB0054

Query Match 48.3%; Score 43; DB 2; Length 338;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 3 RLRSINPTGTVLLQL 18
: : ||||| : |||
Db 87 QIRSHLNPKMTMLTYL 102

RESULT 15

S43295

bone morphogenetic protein homolog GDF6 precursor - mouse (fragment)
 N:Alternate names: growth and differentiation factor 6
 C:Species: Mus musculus (house mouse)
 C:Date: 20-Oct-1994 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
 C:Accession: S43295
 R:Storm, E.E.; Huynh, T.V.; Copeland, N.G.; Jenkins, N.A.; Kingsley, D.M.; Lee, S.J.
 Nature 368, 639-643, 1994
 A:Title: Limb alterations in brachypodism mice due to mutations in a new member of the TGF-beta family
 A:Reference number: S43294; MUID:94195427; PMID:8145850
 A:Accession: S43295
 A:Molecule type: DNA
 A:Residues: 1-125 <STO>
 A:Cross-references: UNIPROT:P43028; UNIPARC:UPI0000024471; EMBL:U08338; NID:9488463; PIR:G488463
 C:Genetics:
 C:Superfamily: inhibin
 F:1-5/Domain: polybasic protease recognition site #status predicted <ppr>
 F:6-125/Product: bone morphogenetic protein homolog GDF6 (fragment) #status predicted <M

Query Match 47.2%; Score 42; DB 2; Length 125;
 Best Local Similarity 50.0%; Pred. No. 10;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 LRSHINPTGVLLQ 17
 ||||: || :||
 Db 61 LRSHLEPTNHAIQ 74

Search completed: December 31, 2005, 23:47:44
 Job time : 3.44828 secs

This Page Blank (uspto)

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	89	100.0	18	5	AAE19258	Human rec	
2	89	100.0	18	7	AD084768	Alpha-gal	
3	89	100.0	18	8	ADJ88294	Tobacco m	
4	89	100.0	18	8	ADM48698	Human rGA	
5	89	100.0	18	8	ADU66933	Human alp	
6	89	100.0	18	9	AEA27462	Human rGA	
7	89	100.0	22	5	AAE19256	Human rec	
8	89	100.0	22	7	AD084766	Alpha-gal	
9	89	100.0	22	8	ADJ88292	Tobacco m	
10	89	100.0	22	8	ADM48696	Human rGA	
11	89	100.0	22	8	ADU66931	Human alp	
12	89	100.0	22	9	AEA27460	Human rGA	
13	89	100.0	24	5	AAE19259	Human rec	
14	89	100.0	24	7	AD084769	Alpha-gal	
15	89	100.0	24	8	ADJ88295	Tobacco m	
16	89	100.0	24	8	ADM48699	Human rGA	
17	89	100.0	24	8	ADU66934	Human alp	
18	89	100.0	24	9	AEA27463	Human rGA	
19	89	100.0	25	9	AEA27477	Human rGA	
20	89	100.0	26	5	AAE19254	Human rec	
21	89	100.0	26	7	AD084764	Alpha-gal	
22	89	100.0	26	8	ADJ88290	Tobacco m	
23	89	100.0	26	8	ADM48694	Human rGA	
24	89	100.0	26	8	ADU66929	Human alp	

```
CC galactosidase-A C-terminal modified peptide
XX
SQ Sequence 18 AA;
  Query Match      100.0%; Score 89; DB 5; Length 18;
  Best Local Similarity 100.0%; Pred. No. 5.4e-08;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 2
ADD84768
ID ADD84768 standard; peptide; 18 AA.
XX
AC ADD84768;
XX
DT 29-JAN-2004 (first entry)
XX
DE Alpha-galactosidase derivative peptide fragment #7.
XX
KW Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Tobacco mosaic virus.
XX
PN US2003106095-A1.
PD 05-JUN-2003.
PF 20-MAR-2002; 2002US-00103327.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2003-801257/75.
XX
PT New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
PS Example 11; SEQ ID NO 29; 77pp; English.
XX
CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents an alpha-galactosidase derivative
CC peptide used in the method of the invention.
XX
SQ Sequence 18 AA;
  Query Match      100.0%; Score 89; DB 7; Length 18;
  Best Local Similarity 100.0%; Pred. No. 5.4e-08;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 3
ADJ88294
ID ADJ88294 standard; peptide; 18 AA.
XX
AC ADM48698;
XX
DT 06-MAY-2004 (first entry)
XX
DE Tobacco mosaic virus WT rGAL-12 (galactosidase) C-terminal peptide.
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX
OS Tobacco mosaic virus.
XX
PN US2004016021-A1.
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
PR 26-FEB-1988; 88US-00160766.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310861.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (FOGU/) FOGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Fogue GP, Erwin RL, Grill LK;
XX
DR WPI; 2004-108227/11.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Example 11; SEQ ID NO 29; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.
XX
SQ Sequence 18 AA;
  Query Match      100.0%; Score 89; DB 8; Length 18;
  Best Local Similarity 100.0%; Pred. No. 5.4e-08;
  Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
   |||||
Db 1 TSRLRSHINPTGTVLLQL 18
   |||||

RESULT 4
ADM48698
ID ADM48698 standard; peptide; 18 AA.
XX
AC ADM48698;
```

```
XX 03-JUN-2004 (first entry)
XX Human rGAL-12 C-terminal peptide.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
XX Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
XX enzyme.
XX
XX Homo sapiens.
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00927733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX
XX Claim 1; SEQ ID NO 29; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human rGAL-12 C-terminal peptide used in the exemplification of the
XX invention.
XX
XX Sequence 18 AA;
XX
XX Query Match 100.0%; Score 89; DB 8; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TSRLRSHINPTGVLLQL 18
XX |||||
XX 1 TSRLRSHINPTGVLLQL 18
XX
XX RESULT 5
XX ADU65933
XX ID ADU65933 standard; peptide; 18 AA.
XX
XX ADU66933;
XX
```

```
XX 10-FEB-2005 (first entry)
XX Human alpha-galactosidase-A C-terminal peptide #7.
XX
XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
XX enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
XX Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
XX Hurler's syndrome; Hurler-Scheie syndrome; nephrotrophic; human.
XX
XX Homo sapiens.
XX
XX US2004234516-A1.
XX
XX 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
XX enzyme replacement therapy for the treatment of lysosomal storage
XX diseases, such as Fabry's disease.
XX
XX Example 11; SEQ ID NO 29; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase enzymes having a post-translational modification provided
XX by the plant expression system. The invention is useful in enzyme
XX replacement therapy for treating lysosomal storage diseases such as
XX Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase-A C-terminal peptide.
XX
XX Sequence 18 AA;
XX
XX Query Match 100.0%; Score 89; DB 8; Length 18;
XX Best Local Similarity 100.0%; Pred. No. 5.4e-08;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TSRLRSHINPTGVLLQL 18
XX |||||
XX 1 TSRLRSHINPTGVLLQL 18
XX
XX RESULT 6
XX AEA27462
XX ID AEA27462 standard; peptide; 18 AA.
XX
XX AEA27462;
XX
XX 11-AUG-2005 (first entry)
XX
XX Human rGAL-12 C-terminal peptide, SEQ ID NO: 29.
XX
XX Gauchers disease; metabolic; neurological disorder; niemann pick disease;
XX genetic disorder; Fabry disease; metabolic disease; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase.
XX
XX Homo sapiens.
XX
```

XX PN US2005125859-A1.
 XX PD 09-JUN-2005.
 XX PF 08-NOV-2004; 2004US-00984389.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-00993059.
 XX PR 20-MAR-2002; 2002US-00103327.
 XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
 XX PI Garger SJ, Turpen TH, Kumagai MH;
 XX DR WPI; 2005-404004/41.
 XX PT New isolated polypeptides useful for producing lysosomal enzymes in
 XX PT plants to be utilized in enzyme replacement therapy or for the
 XX PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
 XX PT Gaucher's disease.
 XX PS Example 11; SEQ ID NO 29; 88pp; English.
 XX CC The present invention relates to the production of human and animal
 XX CC lysosomal enzymes in plants by a transient plant expression system. The
 XX CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 XX CC galactosidase (Gal) enzymes having a post-translational modification
 XX CC provided by the plant expression system. The invention is useful in
 XX CC enzyme replacement therapy for treating lysosomal storage diseases such
 XX CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
 XX CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 XX CC also useful in researches for developing new approaches to medical
 XX CC treatment of lysosomal storage diseases and in industrial processes
 XX CC involving enzymatic substrate hydrolysis. The present sequence is the
 XX CC human galactosidase (GAL) C-terminal peptide present in the vector rGAL-
 XX CC 12.
 XX SQ Sequence 18 AA;
 Query Match 100.0%; Score 89; DB 9; Length 18;
 Best Local Similarity 100.0%; Pred. No. 5.4e-08;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TSLRSHINPTGTVLLQL 18
 Db 1 TSLRSHINPTGTVLLQL 18
 RESULT 7
 AAE19256
 ID AAE19256 standard; peptide; 22 AA.
 AC AAE19256;
 XX 21-MAY-2002 (first entry)
 XX Human recombinant alpha-galactosidase A modified peptide, rGal-8.
 XX Glucocerebrosidase, alpha-galactosidase; Tay-Sachs disease; plant;
 XX enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
 XX Fabry disease; lysosomal storage disease; human.
 XX Homo sapiens.
 XX Synthetic.
 XX WO200208404-A2.
 XX 31-JAN-2002.
 XX 26-JUL-2001; 2001WO-0024111.
 XX 26-JUL-2000; 2000US-00626127.

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.
 XX PI Garger SJ, Turpen TH, Kumagai MH;
 XX DR WPI; 2002-195873/25.
 XX PT New glucocerebrosidase and alpha-galactosidase having a post-
 XX PT transnational modification, useful in enzyme replacement therapy for
 XX PT treating lysosomal storage diseases, e.g. Gaucher disease, Niemann-Pick
 XX PT disease, Fabry disease.
 XX PS Example 11; Fig 5; 102pp; English.
 XX CC The invention relates to production of lysosomal enzymes in plants by
 XX CC transient expression. The invention particularly relates to
 XX CC glucocerebrosidase and alpha-galactosidase having a post-translational
 XX CC modification. The enzymes are useful in enzyme replacement therapy for
 XX CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
 XX CC disease, Fabry disease and Tay-Sachs disease), in researches for
 XX CC developing new approaches to medical treatment of lysosomal storage
 XX CC diseases and in industrial processes involving enzymatic substrate
 XX CC hydrolysis. The present sequence is human recombinant alpha-
 XX CC galactosidase-A C-terminal modified peptide
 XX SQ Sequence 22 AA;
 Query Match 100.0%; Score 89; DB 5; Length 22;
 Best Local Similarity 100.0%; Pred. No. 6.8e-08;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 TSLRSHINPTGTVLLQL 18
 Db 1 TSLRSHINPTGTVLLQL 18
 RESULT 8
 ADD84766
 ID ADD84766 standard; peptide; 22 AA.
 XX 29-JAN-2004 (first entry)
 XX Alpha-galactosidase derivative peptide fragment #5.
 XX Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
 XX enzyme replacement therapy; lysosomal disease; enzyme.
 XX Tobacco mosaic virus.
 XX US2003106095-A1.
 XX 05-JUN-2003.
 XX 20-MAR-2002; 2002US-00103327.
 XX 26-JUL-2000; 2000US-00626127.
 XX 13-NOV-2001; 2001US-00993059.
 XX (GARG/) GARGER S J.
 XX (TURP/) TURPEN T H.
 XX (KUMA/) KUMAGAI M H.
 XX Garger SJ, Turpen TH, Kumagai MH;
 XX WPI; 2003-801257/75.
 XX New polynucleotide for producing active recombinant human and animal
 XX PT lysosomal enzymes in a plant expression system that can be used in enzyme
 XX PT replacement therapy.
 XX PS Example 11; SEQ ID NO 27; 77pp; English.

XX CC The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents an alpha-galactosidase derivative
CC peptide used in the method of the invention.
XX SQ Sequence 22 AA;
Query Match 100.0%; Score 89; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
DB 1 TSRLRSHINPTGTVLLQL 18
RESULT 9
ADJ88292
ID ADJ88292 standard; peptide; 22 AA.
XX AC ADJ88292;
XX DT 06-MAY-2004 (first entry)
XX DE Tobacco mosaic virus WT rGAL-8 (galactosidase) C-terminal peptide.
XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX OS Tobacco mosaic virus.
XX PN US2004016021-A1.
XX PD 22-JAN-2004.
XX PF 23-JUN-2003; 2003US-00602219.
XX PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX DR WPI; 2004-108227/11.
XX PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX PS Example 11; SEQ ID NO 27; 71pp; English.

XX CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's disease. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.
XX SQ Sequence 22 AA;
Query Match 100.0%; Score 89; DB 8; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TSRLRSHINPTGTVLLQL 18
DB 1 TSRLRSHINPTGTVLLQL 18
RESULT 10
ADM48696
ID ADM48696 standard; peptide; 22 AA.
XX AC ADM48696;
XX DT 03-JUN-2004 (first entry)
XX DE Human rGAL-8 C-terminal peptide.
XX KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW enzyme.
XX OS Homo sapiens.
XX PN US2004023281-A1.
XX PD 05-FEB-2004.
XX PF 23-JUN-2003; 2003US-00602220.
XX PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 17-FEB-1989; 89US-00310881.
PR 22-OCT-1990; 90US-00600244.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX DR WPI; 2004-142650/14.
XX PT New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX PS Example 11; SEQ ID NO 27; 72pp; English.
XX CC The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12, rGAL-12R, rGAL-25 or rGAL-25R. The methods and compositions

CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human rGAL-8 C-terminal peptide used in the exemplification of the
CC invention.

XX
XX
SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 8; Length 22;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 11
ADU66931
ID ADU66931 standard; peptide; 22 AA.

XX AC
XX DT 10-FEB-2005 (first entry)
XX DE Human alpha-galactosidase-A C-terminal peptide #5.
XX KW Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
XX KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
XX KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
XX KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.

XX OS Homo sapiens.
XX PN US2004234516-A1.
XX PD 25-NOV-2004.
XX PF 21-MAY-2004; 2004US-00851389.

XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX PR 20-MAR-2002; 2002US-00103327.
XX PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX PI Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX DR

XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.

XX PS Example 11; SEQ ID NO 27; 88pp; English.

XX CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase-A C-terminal peptide.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 8; Length 22;

Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 12
AEA27460
ID AEA27460 standard; peptide; 22 AA.

XX AC AEA27460;

XX DT 11-AUG-2005 (first entry)

XX DE Human rGAL-8 C-terminal peptide, SEQ ID NO: 27.

XX KW Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase.

XX OS Homo sapiens.

XX PN US2005125859-A1.

XX PD 09-JUN-2005.

XX PF 08-NOV-2004; 2004US-00984389.

XX PR 26-JUL-2000; 2000US-00626127.

XX PR 13-NOV-2001; 2001US-00993059.

XX PR 20-MAR-2002; 2002US-00103327.

XX PA (LARG-) LARGE SCALE BIOLOGY CORP.

XX PI Garger SJ, Turpen TH, Kumagai MH;

XX WPI; 2005-404004/41.

XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.

XX PS Claim 22; SEQ ID NO 27; 88pp; English.

XX CC The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human galactosidase (GAL) C-terminal peptide present in the vector rGAL-
CC 8.

XX SQ Sequence 22 AA;

Query Match 100.0%; Score 89; DB 9; Length 22;

Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
DB 1 TSRLRSHINPTGVLLQL 18

RESULT 13


```
AAE19259
ID AAE19259 standard; peptide; 24 AA.
XX
AC AAE19259;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human recombinant alpha-galactosidase A modified peptide, rGal-12R.
XX
KW Glucocerebrosidase; alpha-galactosidase; Tay-Sachs disease; plant;
KW enzyme replacement therapy; Niemann-Pick disease; Gaucher disease;
KW Fabry disease; lysosomal storage disease; human.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200208404-A2.
XX
PD 31-JAN-2002.
XX
PF 26-JUL-2001; 2001WO-US024111.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (LARG-) LARGE SCALE BIOLOGY CORP.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-195873/25.
XX
PT New glucocerebrosidase and alpha-galactosidase having a post-
PT translational modification, useful in enzyme replacement therapy for
PT treating lysosomal storage diseases, e.g. Gaucher disease, Niemann-Pick
PT disease, Fabry disease.
XX
PS Example 11; Fig 5; 102pp; English.
XX
CC The invention relates to production of lysosomal enzymes in plants by
CC transient expression. The invention particularly relates to
CC glucocerebrosidase and alpha-galactosidase having a post-translational
CC modification. The enzymes are useful in enzyme replacement therapy for
CC treating lysosomal storage diseases (e.g. Gaucher disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease), in researches for
CC developing new approaches to medical treatment of lysosomal storage
CC diseases and in industrial processes involving enzymatic substrate
CC hydrolysis. The present sequence is human recombinant alpha-
CC galactosidase-A C-terminal modified peptide
XX
SQ Sequence 24 AA;

Query Match 100.0%; Score 89; DB 5; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRLSHINPTGVLLQL 18
DB 1 TSLRLSHINPTGVLLQL 18

RESULT 14
ADD84769
ID ADD84769 standard; peptide; 24 AA.
XX
AC ADD84769;
XX
DT 29-JAN-2004 (first entry)
XX
DE Alpha-galactosidase derivative peptide fragment #8.
XX
KW Tobacco mosaic virus; alpha-galactosidase; lysosomal enzyme;
KW enzyme replacement therapy; lysosomal disease; enzyme.
XX
OS Tobacco mosaic virus.

Query Match 100.0%; Score 89; DB 5; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSLRLSHINPTGVLLQL 18
DB 1 TSLRLSHINPTGVLLQL 18

RESULT 15
ADJ88295
ID ADJ88295 standard; peptide; 24 AA.
XX
AC ADJ88295;
XX
DT 06-MAY-2004 (first entry)
XX
DE Tobacco mosaic virus WT rGAL-12R (galactosidase) C-terminal peptide.
XX
KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
KW Fabry's disease; Gaucher's disease.
XX
OS Tobacco mosaic virus.
XX
PN US2004016021-A1.
XX
PD 22-JAN-2004.
XX
PF 23-JUN-2003; 2003US-00602219.
XX
PR 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
```

PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (TURP/) TURPEN T H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
PI Turpen TH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-108227/11.
XX
PT New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
PS Example 11; SEQ ID NO 30; 71pp; English.
XX
CC The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is Tobacco
CC mosaic virus galactosidase C-terminal peptide.
XX
SQ Sequence 24 AA;

Query Match 100.0%; Score 89; DB 8; Length 24;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TSRLRSHINPTGVLLQL 18
| | | | | | | | | | | | | | | | | |
Db 1 TSRLRSHINPTGVLLQL 18

Search completed: December 31, 2005, 23:43:15
Job time : 7.57931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 05:47:17 ; Search time 405.497 Seconds
(without alignments)
539.100 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQRNDELHGCALALRFLA.....EWSRLRSHINPTGTLLQL 417

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Xgapop 10.0	Xgapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 4172979 seqs, 262114271 residues

Total number of hits satisfying chosen parameters: 8345958

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140602_5569/app_query.fasta.1.782
-DB=Published Applications NA_New -QFMT=fastap -SUFFIX=p2n.rnpbn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10602220@cgn 1 1 233 @runat_30122005_140602_5569
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DSLEXT=7

Database :

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq2.*
- 9: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq3.*
- 10: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1355	6	US-10-981-267-26 Sequence 26, Appl
2	2293	100.0	1547	6	US-10-981-267-25 Sequence 25, Appl
3	99	4.3	1515	7	US-11-073-112-5 Sequence 5, Appl
4	95.5	4.2	268685	6	US-10-933-025-22 Sequence 22, Appl
5	94	4.1	1241	6	US-10-750-185-56907 Sequence 56907, A
6	93	4.1	1082144	7	US-11-117-187-211 Sequence 211, App
7	91.5	4.0	2177	7	US-11-073-112-4 Sequence 4, Appl
8	90.5	3.9	2716	6	US-10-750-185-31179 Sequence 31179, A

ALIGNMENTS

RESULT 1

US-10-981-267-26
; Sequence 26, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverly, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 26
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Homo sapiens

Sequence 35182, A
Sequence 36, Appl
Sequence 35, Appl
Sequence 16, Appl
Sequence 19, Appl
Sequence 83, Appl
Sequence 56545, A
Sequence 55673, A
Sequence 14, Appl
Sequence 41, Appl
Sequence 215, App
Sequence 1, Appl
Sequence 45064, A
Sequence 58314, A
Sequence 37410, A
Sequence 8553, Ap
Sequence 4, Appl
Sequence 5, Appl
Sequence 6, Appl
Sequence 161, App
Sequence 35565, A
Sequence 45595, A
Sequence 54, Appl
Sequence 107, App
Sequence 1, Appl
Sequence 31, Appl
Sequence 8371, Ap
Sequence 61474, A
Sequence 1, Appl
Sequence 54380, A
Sequence 13307, A
Sequence 13359, A
Sequence 363, App

US-10-750-185-35182
US-11-112-908-36
US-11-112-908-35
US-11-038-284-16
US-11-038-284-19
US-11-108-528-83
US-10-750-185-56545
US-10-750-185-55673
US-11-073-626-4
US-11-038-284-14
US-11-038-284-41
US-10-510-386-235
US-11-055-822-1
US-10-750-185-45064
US-11-073-626-2
US-10-750-185-58314
US-10-750-185-37410
US-10-467-657-8553
US-10-374-954-4
US-10-630-203-5
US-11-103-037-6
US-11-102-240-161
US-10-750-185-35565
US-10-750-185-45595
US-11-121-086-54
US-11-074-176-107
US-11-121-086-3
US-10-467-657-8371
US-10-750-185-61474
US-10-864-758-1
US-10-750-185-54380
US-10-995-561-13307
US-10-995-561-13359
US-10-454-437-363

89.5 3.9 1766 6
88 3.8 98345 7
88 3.8 127340 7
87.5 3.8 2576 9
87.5 3.8 2578 9
87.5 3.8 5301 7
86.5 3.8 1615 6
85 3.7 1212 6
85 3.7 1314 9
84.5 3.7 3033 9
84.5 3.7 3033 9
84 3.7 2866 6
83.5 3.6 948 7
83.5 3.6 1262 6
82 3.6 1314 9
82 3.6 2626 6
81.5 3.6 889 6
81.5 3.6 966 6
81.5 3.6 1335 6
81.5 3.6 1548 6
81.5 3.6 1548 6
81.5 3.6 2380 7
81.5 3.6 2399 6
81.5 3.6 2498 6
81.5 3.6 189252 7
81 3.5 930 7
81 3.5 4860 6
81 3.5 168516 7
80.5 3.5 1242 6
80.5 3.5 1242 6
80.5 3.5 6423 6
80 3.5 1559 6
80 3.5 1611 6
80 3.5 5745 6
80 3.5 35344 6
80 3.5 60844 6
79.5 3.5 867 6

US-10-981-267-26

Alignment Scores:
Pred. No.: 4,55e-261 Length: 1355
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x US-10-981-267-26 (1-1355)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 48 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 107
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 108 CTGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGCAAGCAGCC 167
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 168 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 227
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 228 GATTCTGTCATCAGTGAAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 287
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 288 TGAAGAGTCAAGTATAGTACCTCTGATGATGACTTGTGGATGGCTCCCAAGA 347
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 348 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGATTCGCCAGCTA 407
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 408 GCTAAATATGTTACAGCAAGGACTGAAGCTAGGAGTATTCAGATGTTGGAAATAAA 467
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
DB 468 ACCTGCCAGCGCTTCCCTGGAGTTTGGATACTACAGCAATGATGCCAGACTTGTCT 527
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 528 GACTGGGGAGTAGACTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAAATTG 587
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 588 GCAGATGGTTATAAGCACAATGCTTGGCCCTGAATAGGACTGGCAGAACATGTGTAC 647
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 648 TCCTGTGAGTGGCTCTTATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 707
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 708 CAGTACTGCAATCACTGCGGAAATTTTGTCTGACATTCATGATTCCTGGAAAAAGTATAAG 767
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 768 AGTATCTTGGACTGCATCTTTTAAACAGAGAGAAATGTTGATGTGCTGGACCAAGG 827
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 828 GGTGGATGACCCAGATATGTTAGTGTGGCAACTTGGCCTCAGCTGGAAATCAGCAA 887
QY 281 ValThrGlnMetAlaLeuTrpAlaIleValMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 888 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTTATTCATGTCTTAATGACCTC 947
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320

DB 948 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGCTAAATTGCCATCAAT 1007
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 1008 CAGGACCCCTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGACAACCTTTGAAGTGTGG 1067
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1068 GAACGACCTCTCTCAGGCTTAGCCTGGGTGTAGCTATGATTAACCGGCAGGAGATTGGT 1127
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaLysAsnPro 380
DB 1128 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGTAAAGAGCTGGCCTCTAATCTCT 1187
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1188 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGATGAGACT 1247
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1248 TCAAGGTTAAGAGTCACTAATAATCCACAGGCACCTGTTTGGCTTCAGCTA 1298

RESULT 2

US-10-981-267-25
; Sequence 25, Application US/10981267
; Publication No. US20050281805A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan H
; APPLICANT: Beverley, Stephen
; APPLICANT: Sly, William S.
; TITLE OF INVENTION: TARGETED THERAPEUTIC PROTEINS
; FILE REFERENCE: SYM-009CP
; CURRENT APPLICATION NUMBER: US/10/981,267
; PRIOR FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US 60/384,452
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: US 60/516,990
; PRIOR FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US03/17211
; PRIOR FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US 10/272,531
; PRIOR FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US 60/445,734
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/386,019
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/408,816
; PRIOR FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 25
; LENGTH: 1547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: An exemplary GILT-tagged alpha-GAL A cassette sequence
US-10-981-267-25

Alignment Scores:
Pred. No.: 5,53e-261 Length: 1547
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x US-10-981-267-25 (1-1547)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 51 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 110

```
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 111 CTCGTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGCGCCT 170
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 171 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAGAGCCCA 230
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 231 GATTTCCTGCATCAGTGAGAGAGCTCTCATGTGAGATGGCAGAGCTCATGGTCTCAGAAAGC 290
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 291 TGGAGAGATGCGAGTTATGATGACTCTGCATGTGATGACTGTGGATGGCTCCCCAAAGA 350
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 351 GATTCAAGAGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGGATTCGCCAGCTA 410
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 411 GCTAATTATGTTCACACAAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 470
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 471 ACCTGCGCAGGCTTCCTGGGAGTTTGGATACTACGACATGATGCCAGACCTTTGCT 530
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 531 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAATTTG 590
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 591 GCAGATGGTTATAAGCACATGCTCTTCGCTCTGAATAGGACTGCGCAAGACATTTGTATC 650
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 651 TCCTGTGAGTGGGCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 710
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 711 CAGTACTGCAATCAGTCGCGAAATTTTGTGACATTTGATGATTCCTGGAAAGATATAAG 770
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 771 AGTATCTTGAGTGGACATCTTTTAACGAGAGAGAAATCTTGATGTGCTGGACCCAGG 830
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 831 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 890
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 891 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 950
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 951 CGACATCAGCCCTCAGGCAAAAGCTCTCTCTCAGGATGAAGACGCTAATTCGCATCAAT 1010
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 1011 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1070
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1071 GAACGACCTCTCTCAGGCTTAGGCTGTAGCTATGATGAATAACCGCAGGAGATGGT 1130
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1131 GGAACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGAGTGGCCTGTATCCT 1190
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
```

```
Db 1191 GCTGCTTTCATCATCACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1250
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1251 TCAAGTTAAGAAGTCAATAAATCCACAGGCACTGTTTGTCTTCAAGCTA 1301
RESULT 3
US-11-073-112-5
; Sequence 5, Application US/11073112
; Publication No. US20050260627A1
; GENERAL INFORMATION:
; APPLICANT: Hintz et al.
; TITLE OF INVENTION: Mannosidases and Methods for using the Same
; FILE REFERENCE: 62447-02
; CURRENT APPLICATION NUMBER: US/11/073,112
; PRIOR FILING DATE: 2005-03-03
; PRIOR APPLICATION NUMBER: US 10/089,211
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: PCT/US00/27210
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/157,341
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 5
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Aspergillus nidulans
US-11-073-112-5
Alignment Scores:
Pred. No.: 0.484 Length: 1515
Score: 99.00 Matches: 62
Percent Similarity: 35.74% Conservative: 32
Best Local Similarity: 23.57% Mismatches: 74
Query Match: 4.32% Indels: 95
DB: Gaps: 18
US-10-602-220-16 (1-417) x US-11-073-112-5 (1-1515)
QY 27 ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHis 46
Db 669 CCCGAGCCAGCCAGCACTA-----CGAACCGTTCCTCGGATTTGAT 707
QY 47 TrpGluArgPheMet---CysAsnLeuAspCysGlnGluGluProAspSerCysIleSer 65
Db 708 TGAAGCGCAGTCAATATTGCCGACGCGCAAGCTCGCCAA-----TGTCTCA 752
QY 66 GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaG1 85
Db 753 CATCAGCTG-----GAATGGTGGCGCAGACTCGTA 782
QY 85 YTrpGluTyrLeu-----CysIleAspAs 93
Db 783 CTACGAGTACCTGATCAAGATGTACGTCTACGATCCGAAACGCTTTGGCCTCTACCGGA 842
QY 93 pCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPh 113
Db 843 CGCTGGGTGGCAGCTGGCGAGTCGAGCATCAACCATCTGGCTTCGCACCCGCCACCCG 902
QY 113 eProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGlyI1 133
Db 903 CCCA---GACGTGACTTTCTTGCCCACTTAC-----AACGAGGAGCATCAGCTGGGCGCT 953
QY 133 eTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTy 151
Db 954 GACCAGCCAA-----CACCTGACCTGCTTCGAC-----GGTGGAGAGCTTTCTGCTGTGG 1004
QY 151 rTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLysPheAspG1 171
Db 1005 GACATTGCTGGACCGCCAGGACTTTGTGCACTTCGCGCTTGTGCTGCC-----GG 1058
```

```
QY 171 yCys-----TyrCysAspSerLeuGluAsnLeuAla----- 181
Db 1059 GTGCCACGAGAGGTACAACCTCGACTCTGACGGGCATCGGCCCTGAGCAATTCAGCTGGGA 1118
QY 182 -----AspGlyTyrIlyshiMetSerLeuAlaLeuAsnArgThrGly-- 195
Db 1119 CCTAACGGTGTGCCGACAGCCAGAGAGCTG-----TTCGACGGCGCAGGCTT 1169
QY 196 -----ArgSerIleValTyrSe 201
Db 1170 CTACATCAACAGCGGCCAATACATTCTCGTCCGGAAGTCATCGAGAGCTTCTACTATGC 1229
QY 201 r-----CysGluTrpProLeuTyrMetTrpPropheG1 212
Db 1230 ATGGCGCGTCACAGGTGATGGAACGTACCTCGAATGG-----GTGTGG----- 1272
QY 212 nlyAsProAsnTyrThrGluIleArgGlnTyrCysAsnHisTrpArgAsnPheAlaAsp11 232
Db 1273 ----AAGCGCTTCACCAACATCAACAAGTACTGCCCACTGCGACCGGTTTCGGCGGGCT 1328
QY 232 eaAspAep 234
Db 1329 GGAGAAC 1335

RESULT 4
US-10-933-025-22/c
; Sequence 22, Application US/10933025
; Publication No. US20050265987A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN, STEVEN
; APPLICANT: HEMMERICH, STEFAN
; APPLICANT: TOMITA, MEGUMI
; TITLE OF INVENTION: Sulfotransferases and methods of use
; TITLE OF INVENTION: thereof
; FILE REFERENCE: UCAL-230CON
; CURRENT APPLICATION NUMBER: US/10/933,025
; CURRENT FILING DATE: 2004-09-01
; PRIOR APPLICATION NUMBER: 10/025,966
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/258,577
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 60/267,831
; PRIOR FILING DATE: 2001-09-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 268685
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(268685)
; OTHER INFORMATION: n = A,T,C or G
US-10-933-025-22

Alignment Scores:
Pred. No.: 2,48e+03 Length: 268685
Score: 95.50 Matches: 87
Percent Similarity: 30.73% Conservative: 39
Best Local Similarity: 21.22% Mismatches: 121
Query Match: 4.16% Indels: 163
DB: 6 Gaps: 22

US-10-602-220-16 (1-417) x US-10-933-025-22 (1-268685)
QY 80 GlyTrpLysAspAlaGlyTyrGluTyr---LeuCysIleAspAspCysTrpMet----- 96
Db 57945 GGGTGGGGGATGGAGATGGGAATGGGGTGTAGCAGGAGGTCACTGGGTGTGAGGA 57886
QY 97 -----AlaProGlnArgAspSerGluGly-----Arg 105
Db 57885 GGCCACAGCTGGCCCTCGATGGCCACAGAGTCAGGTATGTGAGGCTCACTGTTCAGAGG 57826
```

```
QY 106 LeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHis 125
Db 57825 AGTGAAGACACCCCAAGA-----ACACAT 57799
QY 126 SerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGlyPhe 145
Db 57798 AGTAAG-----GCTAACACAGGGGAGAGAGGAGGCCAGGGGCTTT 57760
QY 146 ProGlySerPheGlyTyr-----TyrAspIleAspAlaGlnThrPheAlaAspTrp 162
Db 57759 ---GACTCTTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATT 57703
QY 163 GlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeuAlaAsp 182
Db 57702 GGAGTG-----CAGGGGTGTGATCTCGGCTCCCTGCAACCTCCGACTTCC 57658
QY 183 GlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCys 202
Db 57657 TGGTTTCAGCAAAATTCCTGCCTCAGCCTC-----CCAAGT 57625
QY 203 GluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyr 222
Db 57624 AGCTGGGATTACAGGCACAGCCACCCAGCTAAT-----TTT 57583
QY 223 CysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIle 242
Db 57582 TGTATTTTATTAGATGAT-----GGGGTTTGGCCATGTTGG----- 57547
QY 243 LeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrp 262
Db 57546 -----CCAGGCTGGTCT 57535
QY 263 AsnAspProAspMetLeuValIleGlyAsnPheGlyLeu----- 275
Db 57534 CAAACTCCTGACCTCAAGTGATCTGCCACCTCGGCTCCCAAGTACTCGGGGTACAGG 57475
QY 276 -----SerTrpAsnGlnGlnValThrGln 283
Db 57474 CGTGAGCCACCGCCCGGCTCTACTCTTTGTTCCACCTTGGCGCAGCTCCCTGAGCCCT 57415
QY 284 MetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIle 303
Db 57414 GTACAACCTTGGACAGAGGCTGGGGCCAGCTTGTCTCTG-----ATC 57373
QY 304 SerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspPro 323
Db 57372 AGCACCCAGGGCGGCTCTCTGCTGGGCTCTGCTGGAGCATTTCTTG-----CCT 57322
QY 324 LeuGlyLysGlnGlyTyrGlnLeu---ArgGlnGlyAspAsnPheGluValTrp----- 340
Db 57321 GCCTTCCAGCAAGACCTTGGCTTGGCCAGGTGAGTCTGACCTGAGCAGTGTCTTGTCTCG 57262
QY 341 -----GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGlu 358
Db 57261 TCACAGCCTGGTCTCTCTGCCAGGACTTGTATGTAATCGCTGCCACTTG----- 57214
QY 359 IleGlyGlyProArgSerTyr----- 365
Db 57213 -----GCACCCAGGTCTGTCTCACCTCGGATGGCTCTTGGGTGACTCATCAATTGT 57160
QY 366 -----ThrIleAlaValAlaSerLeuGly-----LysGlyValAlaCys 378
Db 57159 GCTTATATCCGGTCTGTGTGGCTCTGCGAGGTCTGTGATGTAATCGGCTGCC 57100
QY 379 AsnProAlaCysPhe----- 383
Db 57099 AACCCAGGCTTGCACAGAGGCTCTGGGGTGAGCCTGTGGGTATCTAGTCGGTTCCTTT 57040
QY 384 -----IleThrGlnLeuLeuProValLysArgLysLeuGly 395
Db 57039 CACCTCAGAGCGAGTGAATTTATGAGTATCACAGCTACTGTGATTGAATATTCACCTGCAC 56980
QY 396 PheTyrGluTrpThrSerArgLeuArgSer 405
```

```

Db      56979 ATATAT---GGGACGTGAGTCTGAGTAGC 56953
RESULT 5
US-10-750-185-56907
; Sequence 56907, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; FILE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; TITLE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56907
; LENGTH: 1241
; TYPE: DNA
; ORGANISM: Bovine 19866881771060
US-10-750-185-56907

Alignment Scores:
Pred. No.: 1.41 Length: 1241
Score: 94.00 Matches: 92
Percent Similarity: 30.51% Conservative: 45
Best Local Similarity: 20.49% Mismatches: 132
Query Match: 4.10% Indels: 180
DB: 6 Gaps: 21

US-10-602-220-16 (1-417) x US-10-750-185-56907 (1-1241)
QY      6 ProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrpAsp 25
      226 CCATTTCTGGAATATCTCTGCTACATGTTCTTTCCACATTTCCGAGGCAAGTGGGAA 285
QY      26 IleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeu 45
      286 -----ACC 288
QY      46 HisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSer-CysIle-- 64
      289 CACTGTGATGACCAATGCTCTCTCCAAAGTCCAGGAGAGGTGGGCAAAACGGTGA 348
QY      65 -----SerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys 82
      349 GCCTGGGAACGGGTGTGTTCTGCGCCCTTGAGAA-----TGGGG 390
QY      82 sAspAlaGlyTyxGluTyxLeuCysIleAspAspCysTrpMetAlaProGlnArgAspse 102
      391 AGATCCA-----GCTCCTAAGCAGCAAG 414
QY      102 rGlu-----Glyx 105
      415 GGAGTTTCTGGTGTCCCTCCCAAGCACTGGGAGGTGGGACAGGTGGAACCTCAA 474
QY      105 gLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyxValHi 125
      475 GTCTCAGGCCCTGGCAGGGAATTTCCCAAGCAGTT---GGCTGATTTGGATACACTCA 531
QY      125 sSerLysGlyLeuLysLeuGlyIleTyxAlaAspValGlyAsnLysThrCysAlaGlyPh 145
      532 TGCTAAGTCCCTC-----TGTGAGCTAGAATTTTTCACITTCCTTTGGGAT 576
QY      145 eProGlySerPheGlyTyxTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyValas 165
      |||||

```

RESULT 6

```

US-11-117-187-211
; Sequence 211, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18

```

; NUMBER OF SEQ ID NOS: 212		; SOFTWARE: Patentin Ver. 2.1	
; SEQ ID NO 211		; LENGTH: 1082144	
; TYPE: DNA		; ORGANISM: Arabidopsis thaliana	
US-11-117-187-211			
Alignment Scores:			
Pred. No.:	3.78e+04	Length:	1082144
Score:	93.00	Matches:	76
Percent Similarity:	30.97%	Conservative:	42
Best Local Similarity:	19.95%	Mismatches:	117
Query Match:	4.06%	Indels:	146
DB:	7	Gaps:	20
US-10-602-220-16 (1-417) x US-11-117-187-211 (1-1082144)			
QY	3	LeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuVal	22
DB	565984	CTGCGTTATCCGAGGTTAGA---GCGGCGCGCGCGCTTAGAATAGTCGGA	566034
QY	23	SerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMet	42
DB	566035	-----GATCTTCGGAACCTCGGTTTGCTGAT-----TTCACACTCGCATG	566076
QY	43	GlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSer	62
DB	566077	GACTTGATGGAAGCTGGTCTTCTCTCTTTGTTTCAATCGAATACGTTGGAATCAG	566136
QY	63	CysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys	82
DB	566137	-----CGAGAGAATCTGCTTCACTAGCT-----	566163
QY	83	AspAlaGlyTyrrGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSer	102
DB	566164	-----AACTCT	566169
QY	103	GluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly	116
DB	566170	CAAAATCGGCTGCA--CCTCGCCAGCATCCCGCGGTCTCGATCGGACTGTCTC	566226
QY	117	-----IleArgGlnLeuAlaAsnTyrrValHisSerLysGlyLeu	129
DB	566227	CGGAGATTTCCGAAGAAGCTGCTTCTGTAATATACAAATGGTCTGTTCTCGGAGTG	566286
QY	130	LysLeuGlyIleTyrrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe	149
DB	566287	AGGTGCCAGTCAACGACCTTATCCAAAGCCGT-----	566319
QY	150	GlyTyrrTyrrAspIleAspAlaGlnThrPheAla	160
DB	566320	-----CACCAACGAACGCCGTTTGAATCTCCGCGTGAGCTTCTC	566361
QY	161	-----AspTrpGly-----ValAspLeuLeuLysPheAspGly	171
DB	566362	TATGTTGCTCTCTATCTATTGATTTGGGGAGAATCTCGGAATCTTCTGTTTATGCTCTGAG	566421
QY	172	CysTyrrCysAspSerLeuGluAsnLeuAlaAspGlyTyrrLysHisMetSerLeuAlaLeu	191
DB	566422	TGCTGTGTACATT-----TTTCATCATATGGCTATGGAACCTT	566460
QY	192	AsnArg-----ThrGlyArgSerIleValTyrrSerCys	202
DB	566461	AATAAAGTCTCGCGGAGAGTTTGATGATATGACTGGAATGCCTTAC-----	566508
QY	203	GluTrpProLeuTyrr-----MetTrpProPheGln	212
DB	566509	---TGGCCTTCTCTCAGGTGATTTGTTCTTCTCAAGAGTGTGTCTATGCTCTATTAC	566565
QY	213	LysProAsnTyrrThrGluIleArgGlnTyrrCysAsn-----HisTrp	226
DB	566566	AAGACGGTTAAGACGGAGGTTGAGTAGCAACATCGGACGAAGCCACACTCTGCTTGG	566625

QY	227	ArgAsnPheAlaAspIleAspAsp-----SerTrpLysSerIleLys	240
DB	566626	AGAAATTAAGCAGCATCAATGAGTATTTCTGGAGTAAGAGAGCTTGAAGACCTTAA	566685
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	566686	TGGCCTCTTGACTACACCACT-----AATTTCTTTGATACTACA-----	566724
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeu-----	275
DB	566725	-----CCTAAAGCAGTAGAGTGGGGAACAGAGGATTTGTGGAGCAGAGATCA	566772
QY	276	SerTrpAsnGlnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPhe	295
DB	566773	TTTTGGAATGTGTATCGAAGCTTTGATCGGTTATGGATTTTATTCGTT-----CTGTAC	566826
QY	296	MetSerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAsp	315
DB	566827	TTGCAGGACGCTATTATTGTTGCTACTAGTATGTTAAGTTTCTTGGCAAGACCGTGAT	566886
QY	316	Val 316	
DB	566887	GTA 566889	
RESULT 7			
US-11-073-112-4			
; Sequence 4, Application US/11073112			
; Publication No. US20050260627A1			
; GENERAL INFORMATION:			
; APPLICANT: Hintz et al.			
; TITLE OF INVENTION: Mannosidases and Methods for using the Same			
; FILE REFERENCE: 62447-02			
; CURRENT APPLICATION NUMBER: US/11/073,112			
; CURRENT FILING DATE: 2005-03-03			
; PRIOR APPLICATION NUMBER: US 10/089,211			
; PRIOR FILING DATE: 2002-02-25			
; PRIOR APPLICATION NUMBER: PCT/US00/27210			
; PRIOR FILING DATE: 2000-10-02			
; PRIOR APPLICATION NUMBER: 60/157,341			
; PRIOR FILING DATE: 1999-10-01			
; NUMBER OF SEQ ID NOS: 25			
; SOFTWARE: Patentin version 3.3			
; SEQ ID NO 4			
; LENGTH: 2177			
; TYPE: DNA			
; ORGANISM: Aspergillus nidulans			
US-11-073-112-4			
Alignment Scores:			
Pred. No.:	6.37	Length:	2177
Score:	91.50	Matches:	45
Percent Similarity:	41.72%	Conservative:	23
Best Local Similarity:	27.61%	Mismatches:	49
Query Match:	3.99%	Indels:	46
DB:	7	Gaps:	11
US-10-602-220-16 (1-417) x US-11-073-112-4 (1-2177)			
QY	27	ProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHis	46
DB	1059	CCCGAGCCAGCCAGTGA-----CGAACCGTTCCTCGATTTGAT	1097
QY	47	TrpGluArgPheMet---CysAsnLeuAspCysGlnGluProAspSerCysIleSer	65
DB	1098	TGGAAGCGCATCAATATTGCCGACGCAAGCTCGCCA-----TGGTCA	1142
QY	66	GluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLys-AspAlaG1	85
DB	1143	CATCAGCTG-----GAATGGTGGCGCAGACTCGTA	1172
QY	85	YTyrrGluTyrrLeu-----CysIleAspAs 93	


```
Db 1173 CTACGAGTACCTGATCAAGATGTACGTCTACGATCCGACGCTTTGGCTCTACCGGGA 1232
Qy 93 pCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPh 113
Db 1233 CCCTGGGTGGCAGCTGCCAGTCCGAGTCAACCATCTGGCTTCGCACCCGCCACCGG 1292
Qy 113 eProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGly 133
Db 1293 CCA---GAGTGCATCTTTGGCCACTTAC-----ACGAGGAGCATCAGCTGGGCCCT 1343
Qy 133 eTyrAlaAspValGlyAsnLysThrCysAlaGlyPheProGlySerPhe-----GlyTyr 151
Db 1344 GACCAGCCAA-----CACCTGACCTGCTTCGAC---GGTGAAGCTTTCTGCTTGGTGG 1394
Qy 151 rTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGly 171
Db 1395 GACATTGCTGACCGCCAGGACTTTGTCGACTTTCGGCTTGCACCTTGTGCC-----GG 1448
Qy 171 yCysTyr 173
Db 1449 CTGCCAC 1455

RESULT 8
US-10-750-185-31179
; Sequence 31179, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 31179
; LENGTH: 2716
; TYPE: DNA
; ORGANISM: Bovine 19866880843346
US-10-750-185-31179

Alignment Scores:
Pred. No.: 11.6 Length: 2716
Score: 90.50 Matches: 48
Percent Similarity: 30.84% Conservative: 22
Best Local Similarity: 21.15% Mismatches: 82
Query Match: 3.95% Indels: 76
DB: 6 Gaps: 11

US-10-602-220-16 (1-417) x US-10-750-185-31179 (1-2716)
Qy 38 ArgThrProThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCys--- 56
Db 1674 AGTTCCCGAGCTCGGAGCCCTCAGCTGTAAAGATC-----CACCTGAATCGGG 1727
Qy 57 GlnGluGluProAspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMet 76
Db 1728 AGACCTGAGTTCGATCCCTCGGTGGGAG----- 1757
Qy 77 ValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMet 96
Db 1758 -----ATCCCTCGGAGAGGGGAGGCTCTCCACTCTGGCCCTGGAGAAATTTTCATGGACT 1811
Qy 97 AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly 116
Db 1812 GTATAGTCCATAGGGTCCGAGAGTCGGACACAA-----CCAAGCGACTTT----- 1856
```

```
Qy 117 IleArgGlnLeuAlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAsp 136
Db 1857 -----CACTTCATTCGGGACCAAGGGCTGTATTGGCGTT----- 1892
Qy 137 ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAla 156
Db 1893 -----TGTACCAAGTTCCTCTGTCTTTGGCTGGGCAAGGCGGATGCT 1937
Qy 157 GlnThrPheAlaAspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSer 176
Db 1938 ACCTCTGTTGGCTC-TGG-----TGCAGCTGT 1963
Qy 177 LeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg 196
Db 1964 GTCCAGAGTCTGGCTGATGGA-----GGTGTCTCAGGTGGCAGGTCCAGGACTG 2011
Qy 197 SerIleValTyrSerCys-----GluTyrProLeu 206
Db 2012 GGAGTGGCCAGGGCTGTGGCCAGGGCCCTGACAGCCTCTGCCACTCTCACTGACACACAG 2071
Qy 207 TyrMetTrpProPheGlnLys-----ProAsnTyr 216
Db 2072 GCATGTGTGCCGACCATCTCTAGGAGTGTGCTCTGCCAAGCCCTCTTTGTCCACCGCTC 2131
Qy 217 ThrGluIleArgGlnTyr-----CysAsnHisTrpArgAsnPhe 229
Db 2132 AGCTCTCTCCAGGCCCAAGAGCTGCCACCCCTCTCTCTGCTGCCCTGGGGAGACCTT 2191
Qy 230 AlaAspIleAspAspSerTrp 236
Db 2192 TCCCACTGGGACCAAGTCATGG 2212

RESULT 9
US-10-750-185-35182
; Sequence 35182, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 35182
; LENGTH: 1766
; TYPE: DNA
; ORGANISM: Bovine 19866880928175
US-10-750-185-35182

Alignment Scores:
Pred. No.: 8.09 Length: 1766
Score: 89.50 Matches: 57
Percent Similarity: 31.33% Conservative: 21
Best Local Similarity: 22.89% Mismatches: 85
Query Match: 3.90% Indels: 86
DB: 6 Gaps: 15

US-10-602-220-16 (1-417) x US-10-750-185-35182 (1-1766)
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAla----- 37
Db 481 CTCTGTCCATGGGATTTCCAGGCAAG-----AATACTGGATCATGTTGCCATTTC 531
```

```
QY 38 -----ArgThrProThrMetGly-----TriLeuHisTrpGlu-Ar 49
Db 532 CTTCTCCAGGAGATCTTGCCAAACCAAGGATCGAACCCATTCTCTTCACTGGGAGCG 591
QY 49 gPheMetCysAenLeuAepCysGlnGluProAspSerCysIleSerGluLysLeuPh 69
Db 592 GATTCCTTCACTGAGCCAGGAGAGCCAGTAGCTGTGTAGCACATTTTATA 651
QY 69 eMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAsp-----AlaG1 85
Db 652 CCTTCAA-----GAGGAAAGGTGGGAGGAGAAACAGTCTTGGGGGCGAG 693
QY 85 YTrpGluTrpLeu-----CysIleAspAsp-----CysTrpMetAlaProGlnArgAspSe 102
Db 694 APATGCCCTCTGCTCTGTTCTTCTCTCTGTTCTGAGGCTGGAGCATCAGG----- 748
QY 102 rGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAs 122
Db 749 -GAGGGATGCTGGTGCCTCT----- 769
QY 122 nTyrValHisSerLysGlyLeuLysGlyIleTyrAlaAspValGlyAsnLysThrCy 142
Db 770 -----GATGTCAAGAGCCCAAGCTTGCCCACTTAAGTCAGGGGAACTGCAGACTG 822
QY 142 salaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTr 162
Db 823 CCAGGGGCTGACAGAGTCAATCGGA-----GGCTG 852
QY 162 pGlyValAspLeuLeuLys-----PheAspGlyCysTyrCysAs 175
Db 853 GAGAACTGGGCTTACAAAGACGAGGAGCAATCTGAGGCTGGGATGCACAAATGGGACAG 912
QY 175 pSerLeu-----GluAsnLeuAlaAspG1 183
Db 913 TGCTCTAGGGAAGCAGGCTGGGTAGTGGGTGACTTCTTAGGAGAGGGGCCACTGG 972
QY 183 YTrpLysHisMetSerLeuAlaLeuAsnArgThr-----GlyArgSerIleVa 199
Db 973 CTATGTCCAGTGTGCATCACCGACCCAGGATTTTAGTTCCAGGGGGAGAGATCCAGT 1032
QY 199 lTyrSerCys---GluTrpProLeuTyrMetTrpProPheGlnLysPAsnTyrThrG1 218
Db 1033 TGCTTCTGTGTCTCTGGGTGTC-----CCTTCTCTCTACCA 1071
QY 218 uIleArgGlnTyrCysAsnHisTrp 226
Db 1072 GACTGAACCTGTAGGAGGACACTGG 1096

RESULT 10
US-11-112-908-36
; Sequence 36, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 36
; LENGTH: 98345
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (49350)..(49350)
; OTHER INFORMATION: n is a, c, g, or t
US-11-112-908-36

Alignment Scores:
Pred. No.: 4,4e+03 Length: 98345
Score: 88.00 Matches: 49
Percent Similarity: 32.43% Conservatives: 15
Best Local Similarity: 24.81% Mismatches: 75
Query Match: 3.84% Indels: 58
DB: 7 Gaps: 11

US-10-602-220-16 (1-417) x US-11-112-908-36 (1-98345)
QY 225 HisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLysSerIleLeuAsp 244
Db 76964 CATTGGCGA----- 76972
QY 245 TrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAsp 264
Db 76973 -----CAGAGCGGGACACGCGCCCTCGGGCCCGAGAGGG----- 77008
QY 265 ProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
Db 77009 ---GAGCGGTACAGCGGGCGCGTGGCACCGCGGTGGAACTCGGCGGGGGGTCC 77065
QY 283 GlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHis 302
Db 77066 GCCCTCTGGCCCTGGCGCCCGCCCGCTGCTTGTCTTCTCG----- 77110
QY 303 lIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAsp 322
Db 77111 CTCGCCCTCTCCCGCCCGCTCGTTCAGCGCGCGGAGGAGCTAACCGCGGT 77170
QY 323 ProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArg 342
Db 77171 CTTCTAGGAACCTCGGGCGGGCAGCAGCCCGCGGA-----TTCTGCTGGCGTCT 77221
QY 343 ProLeu-----SerGlyLeuAlaTrpAlaValAla-----MetIleAsnArgGlnGluIle 359
Db 77222 CGCTGCCATGAAGCGGGACCGCTGCGCGCTTCTGTCTCTGGGTCTGCCGACAGT 77281
QY 360 GlyGlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsn 379
Db 77282 GCGGGGCGCTCGG-----ACGCGCGCGCGCGGTACGCGGACTCGGGCGCGC----- 77329
QY 380 ProAlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrp 399
Db 77330 -----CTTCCCTTAGCGTGGCGAGGG-----TGG 77356
QY 400 ThrSerArgLeuArg-----SerHisIleAsnProThrGlyThr 412
Db 77357 ACGGGCGAGCGCGCGCGCGCTGGGGCGCGGTCTGGGGTCTCTCGCGGGACG 77407

RESULT 11
US-11-112-908-35
; Sequence 35, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
```

; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 35
; LENGTH: 127340
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-35

Alignment Scores:
Pred. No.: 6.43e+03 Length: 127340
Score: 88.00 Matches: 49
Percent Similarity: 32.49% Conservatives: 15
Best Local Similarity: 24.87% Mismatches: 75
Query Match: 3.84% Indels: 58
DB: 7 Gaps: 11

US-10-602-220-16 (1-417) x US-11-112-908-35 (1-127340)

```
QY 225 HieTrrpAgnPheAlaAspSerTrpLysSerIleLeuAsp 244
|||||
Db 25035 CATGGCGA----- 25043
QY 245 TrpThrSerPheAsnGlnGluArgIleValaspValalaGlyProGlyGlyTrpAsnAsp 264
|||||
Db 25044 -----CAGAGCGGGGACGCCGCCCTCGGGCCCGAGAGGG----- 25079
QY 265 ProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsn-----GlnGlnValThr 282
|||||
Db 25080 ---GAGGGTACAGCGGGGGCGCTGGCAGCGGGTGGAACTCGCGCGGGGGTCCC 25136
QY 283 GlnMetAlaLeuTrpAlaIleMetAlaProLeuPheMetSerAsnAspLeuArgHis 302
|||||
Db 25137 GCCCTGGCCCTGGCCGCCGCCCGCTCGCTTGTCTCTCG----- 25181
QY 303 IleSerProGlnAlaIysAlaLeuGlnAspIysAspValIleAlaIleAsnGlnAsp 322
|||||
Db 25182 CTCGGCCCTCCCGCCCGCGCTCGCTTCCAGCGCGCGAGAGCCTTAACCGCGGGT 25241
QY 323 ProLeuGlyLysGlnGlyTyrrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArg 342
|||||
Db 25242 CCTTAGGACCTCGGGCGGGGAGACCGCGCGGA-----TTCTGCTGGCGTCT 25292
QY 343 ProLeu-----SerGlyLeuAlaTrpAlaValAla-----MetIleAsnArgGlnGluIle 359
|||||
Db 25293 CGCTGCCATGAAGCGGACCGCTGGCGCGCTTCTGCTCTCGTGTCTGCCGACAGT 25352
QY 360 GlyGlyProArgSerTyrrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsn 379
|||||
Db 25353 GCGGGGCGCTCGG-----ACGGCGGGCGGGCGCTCAGCGGACTCGGGGCGGC----- 25400
QY 380 ProAlaCysPheIleThrGlnLeuLeuProValIysArgIysLeuGlyPheTyrrGluTrp 399
|||||
Db 25401 -----CTTCCCTTAGCGGTGGCGCGAGGG-----TGG 25427
QY 400 ThrSerArgLeuArg-----SerHisIleAsnProThrGlyThr 412
|||||
Db 25428 ACGGGCGAGCGGCGCGCGCGCTGGGGCCCGGTGGGGTCTCTCGGGGAGC 25478
```

RESULT 12

US-11-038-284-16
; Sequence 16, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION

; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; PRIOR FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 2576
; TYPE: DNA
; ORGANISM: Solanum tuberosum
US-11-038-284-16

Alignment Scores:
Pred. No.: 24.3 Length: 2576
Score: 87.50 Matches: 43
Percent Similarity: 37.14% Conservatives: 22
Best Local Similarity: 24.57% Mismatches: 49
Query Match: 3.82% Indels: 61
DB: 9 Gaps: 9

US-10-602-220-16 (1-417) x US-11-038-284-16 (1-2576)

```
QY 108 AlaAspProGlnArgPhe-----ProHisGlyIleArgGlnLeuAlaAsnTyrrValHis 125
|||||
Db 1138 GCACCAAGCAGCGCTTTGGACGCCCGACGACCTTAAGTCTTTGATTGTAAGACTCAT 1197
QY 126 SerLysGlyLeuLysLeu----- 131
|||||
Db 1198 GAGCTAGGAATTGTTGTTCTCATGGACATGTTTCACAGCCATGCATCAATAATACTTTA 1257
QY 132 ---GlyIleTyrrAlaAspValGlyAsnLysTyrrCysAlaGlyPheProGlySerPheGly 150
|||||
Db 1258 GATGGACTGAACATGTTTGACGGCAGCATAGTGTACTTTCACTCTGGAGCTCGTGGT 1317
QY 151 TyrrTyrrAspIle---AspAlaGlnThrPhe----- 159
|||||
Db 1318 TATCATGTGATGGGATTCCTCCGCTTTTAACTATGGAACCTGGGAGGTACTTAGGTAT 1377
QY 160 -----AlaAspTrpGlyValaspLeuLysPheAspGlyCysTyrrCysAsp 175
|||||
Db 1378 CTCTCTCAAAATGCGAGATGGTGGTGGATGAGTTCAAATTTGATGGATTAGATTGAT 1437
QY 176 SerLeuGluAsnLeuAlaAspGlyTyrrLysHisMetSerLeuAlaLeuAsnArgThrGly 195
|||||
Db 1438 GGTGTGACATCAATGATG-----TATACTCACCGGATTTATCGGTGGGATTCACCTGG 1491
QY 196 Arg-----SerIleValTyrrSer-CysG1 203
|||||
Db 1492 AACTACAGCAATATCTTTCGACTCGCACTGATGTGGTGTCTGTGTATCTGATCTG 1551
QY 203 utrrProLeu-TyrrMetTrpProPheGlnLysProAsnTyrrThrGluIleArgGlnTyrrC 223
|||||
Db 1552 GTCAAGCATCTTATTCATGCGGCTTTTCCAGATGCATTTAC----- 1592
```

RESULT 13

US-11-038-284-19
; Sequence 19, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN

Qy 197 rtle-----ValTyrSerCysGluTrpPro----- 205
Db 2515 ACTTAATAGATCTTCAATATGCTTGGCTCAACCGTGAAGTTATAGCAGATGACTTGC 2574
Qy 206 -----LeuTyrMetTrpProPheGlnLysProAs 215
Db 2575 CTCATCTCTTGGCTTAACCTAGTACCAAGATATATCTACTGCGAGGACTGGAGCCGAC 2634
Qy 215 nTyrThrGluLeuArgGlnTyrCysAsnHisTrp-ArgAsnPheAlaAspIleAspAsps 235
Db 2635 GCAGCATGAGCGGCCAACAA---ACCAGTGGCCAAACCGCCACCATTCATTGAGGCC 2691
Qy 235 erTPrLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArgIleVala 255
Db 2692 ATTGGATTATGTGATGACATCTCTCTTCACTCATCTCGACATCAGGCGTGAATG 2751
Qy 255 spValAlaGlyProGlyGlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyL 275
Db 2752 AATGTCTCCAGCAATGGGCACTGCTC-CCACCTCTGCTGGCTGGCAGTTGGGGG- 2809
Qy 275 euSerTrpAsnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuP 295
Db 2810 -----TTTGTGTGTG-----ATGCCCTGCCACTAC 2837
Qy 295 he-MetSerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuLeu----- 311
Db 2838 TCTCTTAATGTGACAAACAGGACTTGTAGTGTCTCTACGACTTTCCTGTCTTTCAGTCAA 2897
Qy 312 -----GlnAspLysAspValIle----- 317
Db 2898 AAGAGTGGCCATCAACCGCATGGTGAATGAAACACAGAGCCCGACATCATCTCTCC 2957
Qy 318 -----AlaIleAsnGlnAspProLeuGlyLysGlnGlyTyr 329
Db 2958 ATCCAGAGCTTCGGAATGTCGGGCCATTTGACTATGACCCACTGGACAGCAACTCTAT 3017
Qy 330 GlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSerGlyLeuAlaTrp 349
Db 3018 TGGATTGACTCAGCACAAAC----- 3038
Qy 350 AlaValAlaMetIleAsnArgGlnGluLeuGlyGlyProArgSerTyrThrIleAlaVal 369
Db 3039 -----ATGATCCGAAGGACACAGAAGATGGCAGCCAGGCTTTACTGTGTGTG 3089
Qy 370 AlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysPheIleThrGlnLeuLeuPro 389
Db 3090 AGCTCAGT-----CCGAGTCAGAACCTG---GAATACAAACC 3125
Qy 390 ValLysArgLysLeuGlyPheTyrGlu-----TrpThrSerArgLeuArgSer 405
Db 3126 TATGACCTCAGCATTTGATATTTACAGCGCTACATCTACTGAGCTTGTGAGGCTACCAAT 3185
Qy 406 HisIleAsnProThr-----GlyThrValLeu 414
Db 3186 GTCAATTAATGTGAAGATTAGATGGGAGATCAGTTGGAGTGTGTGCTG 3233

RESULT 15

US-10-750-185-56545/c

; Sequence 56545, Application US/10750185

; Publication No. US20050260603A1

; GENERAL INFORMATION:

; APPLICANT: MMI GENOMICS, INC.

; APPLICANT: DENISE, Sue K.

; APPLICANT: KEER, Richard

; APPLICANT: ROSENFELD, David

; APPLICANT: HOLM, Tom

; APPLICANT: BATES, Stephen

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS

; FILE REFERENCE: MM1100-2

; CURRENT APPLICATION NUMBER: US/10750,185

; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482

; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922

; SOFTWARE: PatentIN version 3.1

; SEQ ID NO 56545

; LENGTH: 1615

; TYPE: DNA

; ORGANISM: Bovine 19866880768439

US-10-750-185-56545

Alignment Scores:

Pred. No.: 16.1 Length: 1615
Score: 86.50 Matches: 94
Percent Similarity: 31.52% Conservative: 51
Best Local Similarity: 20.43% Mismatches: 173
Query Match: 3.77% Indels: 144
DB: 6 Gaps: 21

US-10-602-220-16 (1-417) x US-10-750-185-56545 (1-1615)

Qy 23 SerTrpAspIleProGlyAlaArgAlaLeuAspAsnGly----- 35
Db 1435 TCCTGGTAGCTCCTGGGCTGATTCCACAGGCGAGTGGCTCTGTGGTGGGTGGAATTG 1376
Qy 36 -----LeuAlaArgThrProThrMet 42
Db 1375 CAGGAATGAGCAAAAAATTTCCCAAGTCAGGAGGAAAGCAGAGTTCCCAACTCTC 1316
Qy 43 GlyTrpLeuHisTrp---GluArgPheMetCysAsnLeuAspCysGlnGlu----- 59
Db 1315 AGCCACAGAGGTGGCATGAGAGTTGCTACTCTCTGAAATTTGTGGAAGCAGGGTTCG 1256
Qy 60 -----ProAspSerCysIleSerGluLysLeuPheMetGlu 71
Db 1255 CCATCTCTCTCCGGGCACCTGAGCAATGCATCATCAAGGGGGCTTTTATGCAG 1196
Qy 72 MetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTyrGluTyrLeuCysIle 91
Db 1195 AGATTGATGAGGCTTAGGGGAAAAATTTGCAATTTCTTCT-----TAC 1154
Qy 92 AspAspCysTrpMetAlaProGlnArgAspSerGluLysArgLeuGlnAlaAspPro--- 110
Db 1153 GATGATTATTTATAGGCGAATGGATGATGGGATACGAGGCTAAGGTGCTCCATTA 1094
Qy 111 GlnArgPheProHisGlyIleArgGlnLeuAlaAsnTyrValHisSerLys----- 127
Db 1093 GTGAGAAGCCACAGGAGGAGACCTTGTCTGCTTTCGA-CATTTCAAGAGCGGACTC 1035
Qy 127 ----- 127
Db 1034 CTACAGGAGCTGCTTTTCTTTTAGAAGTTTACCTTTAATGTTCTATTTCATGGCCT 975
Qy 128 -----GlyLeuLysLeuGlyIleTyr 134
Db 974 TGTCTGAGCAGGCTGGGAGGAGTAGAGTAGCGGTGCGGGGGCTTTTGGGGAGACAG 915
Qy 135 AlaAspValGlyAsnLysThrCysAla---GlyPheProGlySerPheGlyTyrTyrAsp 153
Db 914 GCGGGGTGGTCACTCACTGAGTGGCAGCTTGGGGCGCGGCGGAGGCTGGGCGAGCGG 855
Qy 154 IleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeuLysPheAspGlyCysTyr 173
Db 854 AGAGCTGTGTGCACG-----GACTGG----- 834
Qy 174 CysAspSerLeuGlnAsnLeuAlaAspGlyTyrLysHisMet---SerLeuAlaLeuAsn 192
Db 833 -----CAGCCCGCGGCCAGCAGCGCCAGCCACTCCGCGGCTGAGCTGCAGG 786
Qy 193 ArgThrGlyArgSerIleValTyrSerCysGluTrpProLeuTyrMetTrpProPheGln 212
Db 785 CGSCCGGCGCAGCAGT-GTTTACTCTGTCTGTCTGCTGCTGCTGCTGCTGCTGCTG 733
Qy 213 LysProAsnTyrThrGluLeuArgGlnTyrCysAsnHisTrpArgAsnPheAlaAspIle 232

Db 732 ---CCGCGCTCCACCGCAGCAGCAGCCTTTTCAACAATCCTGG----- 694
QY 233 AspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSerPheAsnGlnGluArg 252
Db 693 -----CTCTGGCCTGGCCCAAGGCCCTTTTCTCACTATGGACTCTGGGGCTGGACAGA 640
QY 253 IleValAspValAlaGlyProGlyGlyTrp-----AsnAspProAspMetLe 268
Db 639 ATCCCAAACTTGAAG---CCCCGGGCTGGATTGTCTCCGTCCTCTGTGACTGACTCACT 583
QY 268 uValIleGlyAsnPheGlyLeuSerTrpAsnGln---GlnValThrGlnMetAlaLeuTr 287
Db 582 AGGACCCCGAGCGCTGGGCTCAGTTTCCCATGTGTCAATAATCAGATTGATGATTGG 523
QY 287 palatle-MetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnA 307
Db 522 GCAATGTGTTTGGCAGTAGGAGAGTTTCTTCAATGATGTCCAGTGTCCAGGCCCAAGAT 463
QY 307 laLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProLeuGlyLysG 327
Db 462 ACAAGACA-----GATAGCTCAGGTGGGC 439
QY 327 lnglyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrpGluArgProLeuSer---G 346
Db 438 CTGGTCACACCTGGCTCAAGGG-----CCTCTCTCTCTGGCCTTAC 397
QY 346 lylLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGlyGly-----ProA 363
Db 396 TGCTGCCCTGGAAGCTCCTC-----CAGTTTGGGGCTGGGGAGTCCCT 352
QY 363 rgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnProAlaCysP 383
Db 351 GTGGTCCCCCCTAGCAAGCAGTCAGGCAGGAGGCCACTGGCTGTGTGTGTGTGTG 292
QY 383 heIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThrSer 401
Db 291 TGCTGACCCAGAGGACACCACTGCCCCAGGCTGGGT-----TGGAGCTCG 245

Search completed: January 1, 2006, 09:55:33
Job time : 1264.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 03:35:43 ; Search time 1492.57 Seconds
(without alignments)
2310.327 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNPFLHGLCALRFLA.....EWTSLRSHNPVTGTVLLQL 417

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140601_5538/app_query.fasta_1.782
-DB=Published Applications NA Main -QFMT=fastap -SUFFIX=p2n.rnpbm
-MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1
-MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45 -MODE=LOCAL -OUTFMT=ptc -THR SCORE=ptc
-THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext
-HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602220@cgn 1.1 1364 @runat 30122005_140601_5538 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2293	100.0	1254	3	US-09-993-059-15
2	2293	100.0	1254	5	US-10-103-327-15
3	2293	100.0	1254	7	US-10-602-219-15
4	2293	100.0	1254	7	US-10-602-220-15
5	2293	100.0	1254	8	US-10-851-388-15
6	2293	100.0	1254	3	US-10-984-389-15
7	2293	100.0	1266	3	US-09-993-059-11

8	2293	100.0	1266	5	US-10-103-327-11	Sequence 11, Appl
9	2293	100.0	1266	7	US-10-602-219-11	Sequence 11, Appl
10	2293	100.0	1266	7	US-10-602-220-11	Sequence 11, Appl
11	2293	100.0	1266	8	US-10-851-388-11	Sequence 11, Appl
12	2293	100.0	1266	9	US-10-984-389-11	Sequence 11, Appl
13	2293	100.0	1272	3	US-09-993-059-17	Sequence 17, Appl
14	2293	100.0	1272	5	US-10-103-327-17	Sequence 17, Appl
15	2293	100.0	1272	7	US-10-602-219-17	Sequence 17, Appl
16	2293	100.0	1272	7	US-10-602-220-17	Sequence 17, Appl
17	2293	100.0	1272	8	US-10-851-388-17	Sequence 17, Appl
18	2293	100.0	1272	9	US-10-984-389-17	Sequence 17, Appl
19	2293	100.0	1278	3	US-09-993-059-7	Sequence 7, Appl
20	2293	100.0	1278	5	US-10-103-327-7	Sequence 7, Appl
21	2293	100.0	1278	7	US-10-602-219-7	Sequence 7, Appl
22	2293	100.0	1278	7	US-10-602-220-7	Sequence 7, Appl
23	2293	100.0	1278	8	US-10-851-388-7	Sequence 7, Appl
24	2293	100.0	1278	9	US-10-984-389-7	Sequence 13, Appl
25	2293	100.0	1284	3	US-09-993-059-13	Sequence 13, Appl
26	2293	100.0	1284	5	US-10-103-327-13	Sequence 13, Appl
27	2293	100.0	1284	7	US-10-602-219-13	Sequence 13, Appl
28	2293	100.0	1284	7	US-10-602-220-13	Sequence 13, Appl
29	2293	100.0	1284	8	US-10-851-388-13	Sequence 13, Appl
30	2293	100.0	1284	9	US-10-984-389-13	Sequence 13, Appl
31	2293	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
32	2293	100.0	1290	5	US-10-103-327-3	Sequence 3, Appl
33	2293	100.0	1290	7	US-10-602-219-3	Sequence 3, Appl
34	2293	100.0	1290	7	US-10-602-220-3	Sequence 3, Appl
35	2293	100.0	1290	7	US-10-411-037-67	Sequence 67, Appl
36	2293	100.0	1290	7	US-10-411-026-67	Sequence 67, Appl
37	2293	100.0	1290	7	US-10-410-962-67	Sequence 67, Appl
38	2293	100.0	1290	7	US-10-411-049-67	Sequence 67, Appl
39	2293	100.0	1290	7	US-10-410-930-67	Sequence 67, Appl
40	2293	100.0	1290	7	US-10-410-997-67	Sequence 67, Appl
41	2293	100.0	1290	7	US-10-411-012-67	Sequence 67, Appl
42	2293	100.0	1290	7	US-10-410-913-67	Sequence 67, Appl
43	2293	100.0	1290	8	US-10-851-388-3	Sequence 3, Appl
44	2293	100.0	1290	8	US-10-410-980-67	Sequence 67, Appl
45	2293	100.0	1290	9	US-10-410-897-67	Sequence 67, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; TITLE OF INVENTION: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-15 (1-1254)	
QY	1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY	21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db	61 CTGCTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCGCT 120
QY	41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db	121 ACCATGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY	61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db	181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTTCATGGTCTCAGAGGC 240
QY	81 TrpLysAspAlaGlyTyTrpGluTyLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db	241 TGGAGAGATGCAGGTATAGTACCTCTGCAATTCATGATTCATGCTGGTCCCAAGA 300
QY	101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db	301 GATTTCAGAGGCAGACTTCAGGAGACCTTCAGGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360
US-10-103-327-15	
; Sequence 15, Application US/10103327	
; Publication No. US20030106095A1	
; GENERAL INFORMATION:	
; APPLICANT: GARGER, Stephen A.	
; APPLICANT: TURPEN, Thomas H.	
; APPLICANT: KUMAGAI, Monto H.	
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN	
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION	
; FILE REFERENCE: 008010087CPUS06	
; CURRENT APPLICATION NUMBER: US/10/103,327	
; CURRENT FILING DATE: 2002-03-20	
; PRIOR APPLICATION NUMBER: US/09/993,059	
; PRIOR FILING DATE: 2001-11-13	
; NUMBER OF SEQ ID NOS: 37	
; SOFTWARE: FastSeq for Windows Version 4.0	
; SEQ ID NO 15	
; LENGTH: 1254	
; TYPE: DNA	
; ORGANISM: Homo sapiens	
; FEATURE:	
; NAME/KEY: CDS	
; LOCATION: (1)...(1254)	
US-10-103-327-15	
Alignment Scores:	
Pred. No.:	3,73e-275
Score:	2293.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	5
US-10-602-220-16 (1-417) x US-10-103-327-15 (1-1254)	
QY	1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db	1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
QY	21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db	61 CTGCTTCTCGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGAGCGCT 120
QY	41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db	121 ACCATGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY	61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db	181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTTCATGGTCTCAGAGGC 240
QY	81 TrpLysAspAlaGlyTyTrpGluTyLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db	241 TGGAGAGATGCAGGTATAGTACCTCTGCAATTCATGATTCATGCTGGTCCCAAGA 300
QY	101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db	301 GATTTCAGAGGCAGACTTCAGGAGACCTTCAGGAGCCCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360


```
Qy 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACTGCCAGGCTTCCCTGGAGTTTGGATACAGCATGATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTTACTGTGACATGTTGGAAAATTG 540
Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACATGCTCTGGCCCTGAATAGAGCTGGCAGACATTTGTGTAC 600
Qy 201 SerCysGlnTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGCCCAATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCACTGGCGAAATTTTGTGACATTCATCTCTGGAAAAGTATTAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACATGACATCTTTTAACACGAGAGAAATTTGTATGTTGTGGACAGGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGTGAATGACCCAGATATGTAGTATGTTGGCACTTTGGCCCTCAGCTGGAATCAGAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTCTTTATTCATGCTCATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGACCCCTTGGCAGCAAGGGTACCAGCTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTCTGAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCAACACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTACATATAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 3
US-10-602-219-15
; Sequence 15, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
```

```
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-15
```

```
Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
```

US-10-602-220-16 (1-417) x US-10-602-219-15 (1-1254)

```
Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAACTGATTTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGCATTGATGACTGTTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATCGGATTCGCGACTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTCCACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATPAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATACTACGACATTCATGCCAGACCTTTGCT 480
```

```
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
|
|
|
Db 481 GACTGGGAGTAGACTCTGTAATAATTTGATGGTTGTTACTGTGACAGCTTTGGAAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|
|
|
Db 541 CGAGATGGTTATAAGCACATGCTCTTGGCCCTGTAATAGGACTGGCAGAAGCAATTTGTATC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
|
|
|
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|
|
|
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGCATTTGATGATGATTTTATACAGAAATCCGA 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|
|
|
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGTTGCTGGACCAGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|
|
|
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATTTGGCAACTTTTGGCCCTCAGCTGGAATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|
|
|
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|
|
|
Db 901 CGACATCATGACCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|
|
|
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCGTCTAGACAGGGGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
|
|
|
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCCGGCAGGAGATGGT 1080

QY 361 GlyProAsnSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
|
|
|
Db 1081 GGACCTCGCTCTTATACCATCGGAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCTCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|
|
|
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|
|
|
Db 1201 TCAAGGTTAAGAAGTCACATAAAATCCACAGGCACTGTTTTGTCTCAGCTA 1251

RESULT 4
US-10-602-220-15
; Sequence 15, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; PRIOR FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
```

```
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-15

Alignment Scores:
Pred. No.: 3,73e-275 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 7

US-10-602-220-16 (1-417) x US-10-602-220-15 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|
|
|
Db 1 ATGCAGCTCAGGAAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTGCCTTCTCGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|
|
|
Db 61 CTGCTTCTCTGGGACATCTCTGGGGCTAGAGCACTGGACATGTGATTTGGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|
|
|
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluMetValSerGluGly 80
|
|
|
Db 181 GATTCTCTGCATCATGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|
|
|
Db 241 TGGAGAGATGCAGGTATATGATACCTTCATTTGATGACTTGTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|
|
|
Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATTCGCGACGTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|
|
|
Db 361 GCTAAATTTATGTTTACAGCAAGGAGCTGAACTAGGGATTTATGCAGATGTTGGAAATTA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
|
|
|
Db 421 ACCTGCCAGGCTTCCCTGGGAGATTTTGGATACTACGACATTCATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|
|
|
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGTCTTGGAAAAATTTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|
|
|
Db 541 GCAGATGGTTATAAGCACATGTCTTGGCCCTGTAATAGGACTGGCAGAAAGCATTTGTGTAC 600
```

```
QY 201 SerCysGluTrpProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCCCTCTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATGATGATTTCTTGGAAAAGTATANAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGTATGTGTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAACCGGAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGTAAAGAGATGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCTCCCTGTGAAAGGAGGCTAGGCTTCTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTGTCTTCAGCTA 1251
```

RESULT 5

```
US-10-851-388-15
; Sequence 15, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: KUMAGAI, Tomo H.
; APPLICANT: TURPEN, Stephen A.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851.388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-851-388-15
```

```
Alignment Scores: 3.73e-275 Length: 1254
Pred. No.: 2293.00 Matches: 417
Score:
```

```
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-10-602-220-16 (1-417) x US-10-851-388-15 (1-1254)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCGCTTCGCGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCCTGGGACATCTCCCTGGGCTTAGAGCACTGGACAATGGATGGCAAGGACGCCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTACTGCTCCAGGAGAGACCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCGAGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTTACAGCAAGGAGCTGAAGCTAGGAGATTTATGCAGATGTTGAAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATATCTACGACATTTGATGCTGACAGCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGTAATAATTTGATGGTGTGTACTGTGACAGCTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTCAATAGGACTGGCAGAGCAATGCTGAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATGATGATTTCTTGGAAAAGTATANAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAATTTGTATGTGTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAACCGGAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGTAAAGAGATGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGTCTCATCACAGCTCTCTCCCTGTGAAAGGAGGCTAGGCTTCTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTTGTCTTCAGCTA 1251
```

Db 961 CAGGACCCCTTGGCAGCAAGGTACAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCCGGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCCTGTAAATCTT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCTGCTTCATCACAGCTCTCTCTGTGAAAGAGNAGCTAGGGTCTTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 6

US-10-984-389-15

; Sequence 15, Application US/10984389

; Publication No. US20050125859A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN

; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION

; FILE REFERENCE: 008010087CPUS06

; CURRENT APPLICATION NUMBER: US/10/984,389

; CURRENT FILING DATE: 2004-11-08

; PRIOR APPLICATION NUMBER: US/09/993,059

; PRIOR FILING DATE: 2001-11-13

; NUMBER OF SEQ ID NOS: 37

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 1254

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(1254)

US-10-984-389-15

Alignment Scores:

Pred. No.:	3,73e-275	Length:	1254
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-10-602-220-16 (1-417) x US-10-984-389-15 (1-1254)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGTTTCTTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGNAGGATCGAGGTATAGTACCTCTGCATGACTGTGTGATGGCTCCCCAAAGA 300

RESULT 7

US-09-993-059-11

; Sequence 11, Application US/09993059

; Publication No. US20020088024A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

; APPLICANT: TURPEN, Thomas H.

; APPLICANT: KUMAGAI, Monto H.

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGCAGACTTCAGGAGACCCCTCAGCGCTTCTCTCATGGGATTGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTGCGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGAGTTTGGATATACGACATTCATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGCTTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGGTTATAAGCACATGCTTGGCCCTGATAGGACTGGCAGAAAGCATTTGTATC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTAGTGGCCTCTTTATATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACATGGCGAAATTTTGTGCATTCATGATTCCTCGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTATGCTGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGATTGGCACTTTGGCCCTCAGCTCGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGCTAATGACC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCCGGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTCTGTAATCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGAGGAGCTAGGGTCTTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/09/993,059
 ; CURRENT FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 1266
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1266)
 ; US-09-993-059-11

Alignment Scores:
 Pred. No.: 3,78e-275 Length: 1266
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-11 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGCAATGATTTGGCAAGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAAGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGGATGCAGGTTATGAGTACCTCTGCAITGATGACTGTGGATGGCTCCCCAAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAGAGGAGAGACTTCAGGAGACCTTCAGCGCTTTCCCTATGGGATTCGCCAGCTA 360
 QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTyxAlaAspValGlyAsnLys 140
 DB 361 GCTAATTATGTTACAGCAAGAGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGGCAGGCTTCCTCGGAGTTTGGATATACGATGATGATGATGATGATGATGATGAT 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyxCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTATGATCTGCTAAATTTGATGGTTGTTACTGTGACAGATTTGGAATAATG 540
 QY 181 AlaAspGlyTyxLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyx 200
 DB 541 GCAGATGGTTATAGACATGCTCTGGCCCTGATAGACTGGCAGAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATATGTCGCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
 QY 221 GlnTyxCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCAGTCGGCAAAATTTTGTGATTTGATTCCTGGAAAGATTAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGAGCTGGACATCTTTTAAACAGGAGAGAATTTGTTGATGTTGCTGACCAAGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 DB 781 GGTGGAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGGAGAGATTGGT 1080
 QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTATACCATCGAGTTGCTTCCTGGTAAAGAGCTGGCTGTAAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyxGluTrpThr 400
 DB 1141 GCCTGCTTCATCACACAGCTCTCCTGTGAAAAGGAAGCTAGGGTTCTATGATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 DB 1201 TCAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 8

US-10-103-327-11
 ; Sequence 11, Application US/10103327
 ; Publication No. US20030106095A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARGER, Stephen A.
 ; APPLICANT: TURPEN, Thomas H.
 ; APPLICANT: KUMAGAI, Monto H.
 ; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
 ; FILE REFERENCE: 008010087CPUS06
 ; CURRENT APPLICATION NUMBER: US/10/103,327
 ; PRIOR FILING DATE: 2002-03-20
 ; PRIOR APPLICATION NUMBER: US/09/993,059
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 11
 ; LENGTH: 1266
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1266)
 ; US-10-103-327-11

Alignment Scores:
 Pred. No.: 3,78e-275 Length: 1266
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-11 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlnLeuAlaArgThrPro 40
Db 61 CTCTGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACATGGGCTGGCTGCATCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGCAGACCTCAGCGCTTTCCTCATGGATTGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGCGCTTCCCTGGGAGTTTGGATATCTAGCAGATTCATGCCAGACTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GATCGGGAGTAGATCTGCTAAATTTGATGTTGTACTGTACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGGCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATCTACTGGCGAATTTTGTCTGACATTGATGATTCCTTGAAAAGTATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAACAGCAGAGAGAAATTGTATGTTGCTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCGCTTACAGAGGAGACAACCTTTGAATGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAAGACCTCTCTCAGGCTTAGCCCTGGCTGTAGCTATGATATAAACCCGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380

Db 1081 GGACCTCGCTCTTATACCATCGAGTGTCTTCCCTGGGTAAAGGAGTGGCCTGTATTCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGTTTCATCACACAGCTCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 9
US-10-602-219-11
; Sequence 11, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-219-11

Alignment Scores:
Pred. No.: 3 78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 7

US-10-602-220-16 (1-417) x US-10-602-219-11 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlnLeuAlaArgThrPro 40
Db 61 CTCTGTTCTCGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 121 ACCATGGCTGGCTGCATCTGGAGGCTTCATGTGCAACCTTGACTGCCAGAGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGCTCTTCTGAGAGATGGCAGAGCTCATGTCTCAGAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGCTTATGATGATCTCTGCAATGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAGGAGACTTCAGCAGACCTCAGGCTTCTCATGGATTCGCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACAGCAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGGAGCTTCCCTGGAGATTTGGATACACGACATGATGATGCCAGACCTTGGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTCGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTATAAGCATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTCAAAAGGCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTCGGAAATTTGCTGACATTTGATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGCATCTTTTAAACGAGAGAGATTTGTGATGTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAATGACCCAGATATGTTAGTGAATGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGCTTATCATGGCTCTCTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTTCAGATTAAGGACGTAATGTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGACCCCTTGGCAAGCAAGGATACAGCTTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTTCCCTCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417

Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTGCTTCAGCTA 1251
RESULT 10
US-10-602-220-11
; Sequence 11, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Poque, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: USBC-0087-CE07B
; CURRENT APPLICATION NUMBER: US/10/602,220
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-602-220-11
Alignment Scores:
Pred. No.: 3,78e-275 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 7 Gaps: 0
US-10-602-220-16 (1-417) x US-10-602-220-11 (1-1266)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGCACTACATCTGGGCTGGCGCTGCGCTTCCGCTTCTGGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATGTTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGAGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGCTCTTCTCATGGAGATGGCAGAGCTCATGTCTCAGAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100

```
Db 241 TCGAAGGATGCAAGGTATGAGTACCTCTCATGTGATGCTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGAGAGCTTCAGGACAGACCTCAGCGCTTCTCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACACAGCAAGGAGCTGAAGCTAGGATTTATGACAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCTCCGAGTCTTGGATACTACGACATGATGATGCCAGACCTTGTCT 480
Qy 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTCGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCANATCCTGGCGAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATCTTGATGTGCTGGACCGGG 780
Qy 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTGGGAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCTTCAGCTGGATACAGAA 840
Qy 281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGGTACAGCTTAGACGGAGAACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGCACTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAAACCCGCGAGGATTTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGTGAAAAGGAAGCTAGGGTTCATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGTTTAAAGTTCATATAATCCCAAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 11

US-10-851-388-11

; Sequence 11, Application US/10851388

; Publication No. US20040234516A1

; GENERAL INFORMATION:

; APPLICANT: GARGER, Stephen A.

```
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1) ... (1266)
US-10-851-388-11
```

Alignment Scores:

Pred. No.:	3,78e-275	Length:	1266
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0

US-10-602-220-16 (1-417) x US-10-851-388-11 (1-1266)

```
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGACACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTCTGGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCTGGACATCCCTGGGGCTAGAGCACTCGACAAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTCGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrPheMetAlaProGlnArg 100
Db 241 TGGAAAGGATGCAAGGTTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGAGAGACTTCAGGACAGACCTCAGGGCTTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGATTTATGACAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCTCCGAGTCTTGGATACTACGACATGATGATGCCAGACCTTTGTCT 480
Qy 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTCGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
```



```
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||
DB 661 CAGTACTGCAATCACTGCGGAAATTTTGGCTGACATTGATCTCGGAAAGTATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
|||
DB 721 AGTATCTTGAGCTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTGCTGGACACGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
DB 781 GGTGGAAATGACCCAGATATGTTAGTGTGTCAGTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTTATGTCATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||
DB 901 CGACATCATAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|||
DB 961 CAGGACCCCTTGGCAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
|||
DB 1021 GAACGACCTCTCTCAGGCTTAGCGCTGTAGCTATGATATAAACCGCAGGAGATTGGT 1080
QY 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
|||
DB 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
|||
DB 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAGCTAGGGTTCATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|||
DB 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTA 1251
```

RESULT 12

```
US-10-984-389-11
; Sequence 11, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-984-389-11
```

Alignment Scores:

Pred. No.:	3,78e-275	Length:	1266
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

```
DB: 9 Gaps: 0
US-10-602-220-16 (1-417) x US-10-984-389-11 (1-1266)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||
DB 1 ATGCACTGAGGAACCCAGCAACTACATCTGGGCTTGGCGCTTGGCTTCCGCTTCTGCGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||
DB 61 CTCGTTTCTCTGGGACATCCCTGGGCTTAGAGCACTGGCAATGGATGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|||
DB 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||
DB 181 GATTCTTCGTCATCAGTGAGAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||
DB 241 TGGAAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCCAAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|||
DB 301 GATTTCAGAAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||
DB 361 GCTAATTATGTTTCACAGCAAGGACTGAGCTAGGAGATTTATGCAGATGTTGGAAATATA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
|||
DB 421 ACCTGCGCAGGCTTCCTGGGAGTTTGGATATCTAGCACATTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGCACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||
DB 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
|||
DB 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAAGGCCCAATTATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTTGATTCCTGGAAAAGTATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
|||
DB 721 AGTATCTTGAGCTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTGCTGGACACGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||
DB 781 GGTGGAAATGACCCAGATATGTTAGTGTGTTGGCAACTTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTTATGTCATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
|||
DB 901 CGACATCATAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
|||
DB 961 CAGGACCCCTTGGCAGCAAGGCTACAGCTTAGACAGGAGACAACTTTGAAGTGTG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
```

Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATTAACCGCGAGGAGTTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGSTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251

RESULT 13

US-09-993-059-17
; Sequence 17, Application US/09993059
; Publication No. US20020088024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCES: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1272)
US-09-993-059-17

Alignment Scores:
Pred. No.: 3,81e-275 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-17 (1-1272)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCCTGGGACATCCCTGGGGCTAGACACTGGACAAATGATTTGGCAGAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCATCGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTCTTATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGAGGTATAGTACTCTGCAATGATGACTTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGCAGACTTCAGCGAGACCTTCAGCGCTTTCCTCATGGGATTCGCGAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTACAGCAAGAGACTGAAGCTAGGATTTATGCGATGTTGCAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACGACATTTGATGTCGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTGCTGCTAAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTTATAGCACATGCTTGTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACCTGGCGAAATTTTGTCTGACATTTGATGATTCCTCGGAAAGTATTAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGTGATGTTGCTGACCCAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTTCGAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTCGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGGCTATCATGGCTGCTCTTATTATCATGTCATATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACACAGCTTAGACAGGAGACCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATTAACCCGCGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251

RESULT 14

US-10-103-327-17
; Sequence 17, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCES: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20

; PRIOR APPLICATION NUMBER: US/09/993,059
 ; PRIOR FILING DATE: 2001-11-13
 ; NUMBER OF SEQ ID NOS: 37
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 17
 ; LENGTH: 1272
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)...(1272)
 ; US-10-103-327-17

Alignment Scores:
 Pred. No.: 3,81e-275 Length: 1272
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-17 (1-1272)

```

Qy 1 MetClnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGGTTCTCGGNCATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCTGCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGNAGGATCAGGTTATGATGACTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTCAGGAGAGCCCTCAGCGCTTCTCATGGATTCCGACAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTTACAGCAAGAGCTGAAGCTAGGAGTTTATGAGATGTTGGAATAAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACTTGGCAGCTTCCCTGGAGATTTTGGATATACAGCAATGATGCCAGACCTTGGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGGTTGTTTACTGTGACAGTTGATGGAATAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCAGCATGCTCTTGGCCCTTGAATAGGACTGGCAGAACATTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATGATGATGATGATGATGATGATG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

```

```

Db 721 AGTATCTTGACTGGACATCTTTTAACGAGAGAGAAATTTGATGATGTTGCTGGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTTAGACAGGAGACAACTTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGGCTGGCTGTAGCTATGATAAACCCGCGAGGATGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGACTCTCCCTGTTGANAAGGAGCTAGGTTCTATGATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

```

RESULT 15

US-10-602-219-17

; Sequence 17, Application US/10602219

; Publication No. US20040016021A1

; GENERAL INFORMATION:

; APPLICANT: Large Scale Biology Corporation

; APPLICANT: Turpen, Thomas H.

; APPLICANT: Pogue, Gregory P.

; APPLICANT: Erwin, Robert L.

; APPLICANT: Grill, Laurence K.

; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION

; FILE REFERENCE: LSBC-0087-CP09B

; CURRENT FILING DATE: 2003-06-23

; PRIOR APPLICATION NUMBER: 09/993,059

; PRIOR FILING DATE: 2001-11-13

; PRIOR APPLICATION NUMBER: 09/626,127

; PRIOR FILING DATE: 2000-07-26

; PRIOR APPLICATION NUMBER: 09/316,572

; PRIOR FILING DATE: 1999-05-21

; PRIOR APPLICATION NUMBER: 08/324,003

; PRIOR FILING DATE: 1994-10-14

; PRIOR APPLICATION NUMBER: 08/176,414

; PRIOR FILING DATE: 1993-12-29

; PRIOR APPLICATION NUMBER: 07/997,733

; PRIOR FILING DATE: 1992-12-30

; PRIOR APPLICATION NUMBER: 08/184,237

; PRIOR FILING DATE: 1994-01-19

; PRIOR APPLICATION NUMBER: 07/923,692

; PRIOR FILING DATE: 1992-07-31

; PRIOR APPLICATION NUMBER: 07/600,244

; PRIOR FILING DATE: 1990-10-22

; PRIOR APPLICATION NUMBER: 07/641,617

; PRIOR FILING DATE: 1991-01-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 32

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 17

; LENGTH: 1272

! TYPE: DNA
! ORGANISM: Homo sapiens
US-10-602-219-17

Alignment Scores:

3 81e-275 Length: 1272
Pred. No.: 2293.00 Matches: 417
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 100.00% Gaps: 0
DB: 7

US-10-602-220-16 (1-417) x US-10-602-219-17 (1-1272)

QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCTCGGACATCCCTGGGGCTAGAGCAATGGATTGGCAAGAGCGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGGCTGGTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCTCGTCATGATGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC	240
QY	81	TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATCAGGTTATAGTACCTCTGATGATGACTGTGTGGATGGCTCCCAAGA	300
QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	301	GATTGAGAGCGAGACTTCAGGCGAGCCCTCAGCGCTTTCCTCATGGGATCGCCAGCTA	360
QY	121	AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys	140
Db	361	GCTAATTATGTTACAGCAAAAGGACTGAAGCTAGGAGTTTATGCAGATGTTGGAAATAAA	420
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla	160
Db	421	ACCTGCCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGCCAGAGCCCTTGCT	480
QY	161	AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
Db	481	GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTG	540
QY	181	AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp	200
Db	541	GCAGATGGTTATAAGCACAATGCTTGGCCCTGAATAGGACTGGCAGAACGATTGTGTAC	600
QY	201	SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg	220
Db	601	TCCTGTGAGTGGCTCTTTTATATGTGGCCCTTTCAAAGGCCAAATATACAGAAATCCGA	660
QY	221	GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
Db	661	CAGTACTGCAATCACTGGCGNAATTTTGCTGACATTGATGATTCCTGGAAAAGTATAAG	720
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	721	AGTATCTGGAGTGGACATCTTTAAACGAGAGAGAAATTTGTGATGTTGTGGACACAGG	780
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
Db	781	GGTTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA	840
QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	841	GTAACTCAGATGGGCCCTCTGGGCTATCATGGCTGCTCTTTTATTCATGTCTAATGACCTC	900

QY	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
Db	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGATTAAGGACGTAATTGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGGAGACAACTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
Db	1021	GAACGACCTCTCTCAGGCTTAGCTGGGCTAGCTATCATAAACCGGAGGAGATTGGT	1080
QY	361	GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAAGGAGTGGCTGTATTCCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr	400
Db	1141	GCCTGCTTCATCACACAGCTCCTCCCTGTGMAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
Db	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTGCTCAGCTA	1251

Search completed: January 1, 2006, 09:34:32

Job time : 1520.57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:29:37 ; Search time 339.352 Seconds
(without alignments)
2184.292 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNLHLCALRFLA.....EWSRLRSHNPQTGLLQL 417

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1303057 seqs, 888780828 residues
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140559_5451/app_query.fasta_1.782
-DB=Issued Patents NA -QFMP=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10602220@cgn 1.1 237 @runat_30122005_140559_5451 -NCPUS=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1254	3	US-09-993-059-15
2	2293	100.0	1254	3	US-10-103-327-15
3	2293	100.0	1266	3	US-09-993-059-11
4	2293	100.0	1266	3	US-10-103-327-11
5	2293	100.0	1272	3	US-09-993-059-17
6	2293	100.0	1272	3	US-10-103-327-17
7	2293	100.0	1278	3	US-09-993-059-7
8	2293	100.0	1278	3	US-10-103-327-7
9	2293	100.0	1284	3	US-09-993-059-13

10	2293	100.0	1284	3	US-10-103-327-13	Sequence 13, Appl
11	2293	100.0	1290	3	US-09-993-059-3	Sequence 3, Appl
12	2293	100.0	1290	3	US-10-103-327-3	Sequence 3, Appl
13	2293	100.0	1296	3	US-09-993-059-9	Sequence 9, Appl
14	2293	100.0	1296	3	US-10-103-327-9	Sequence 9, Appl
15	2293	100.0	1308	3	US-09-993-059-5	Sequence 5, Appl
16	2293	100.0	1308	3	US-10-103-327-5	Sequence 5, Appl
17	2293	100.0	1343	3	US-08-928-881-18	Sequence 18, Appl
18	2293	100.0	1343	3	US-09-543-921-18	Sequence 18, Appl
19	2293	100.0	1343	3	US-09-266-014-3	Sequence 3, Appl
20	2293	100.0	1343	3	US-09-491-759-18	Sequence 18, Appl
21	2293	100.0	1393	2	US-07-602-8248-1	Sequence 1, Appl
22	2293	100.0	1393	2	US-07-983-451-1	Sequence 1, Appl
23	2293	100.0	1393	2	US-08-261-577-6	Sequence 6, Appl
24	2293	100.0	8234	3	US-09-626-127-14	Sequence 14, Appl
25	2293	100.0	8234	3	US-09-993-059-34	Sequence 34, Appl
26	2293	100.0	8234	3	US-10-103-327-34	Sequence 34, Appl
27	2231	97.3	1233	3	US-09-993-059-21	Sequence 21, Appl
28	2231	97.3	1233	3	US-10-103-327-21	Sequence 21, Appl
29	2227	97.1	1215	3	US-09-993-059-19	Sequence 19, Appl
30	2227	97.1	1215	3	US-10-103-327-19	Sequence 19, Appl
31	2213	96.5	1304	9	5179023-3	Patent No. 5179023
32	2157	94.1	1214	3	US-09-023-655-1001	Sequence 1001, Ap
33	2132	93.0	11641	3	US-09-626-127-13	Sequence 13, Appl
34	2132	93.0	11641	3	US-09-993-059-33	Sequence 33, Appl
35	2132	93.0	11641	3	US-10-103-327-33	Sequence 33, Appl
36	2131	92.9	1158	3	US-09-176-666-53	Sequence 53, Appl
37	2131	92.9	1161	3	US-09-176-666-52	Sequence 52, Appl
38	2131	92.9	1164	3	US-09-176-666-51	Sequence 51, Appl
39	2131	92.9	1167	3	US-09-176-666-50	Sequence 50, Appl
40	2131	92.9	1170	3	US-09-176-666-49	Sequence 49, Appl
41	2131	92.9	1173	3	US-09-176-666-48	Sequence 48, Appl
42	2131	92.9	1176	3	US-09-176-666-47	Sequence 47, Appl
43	2131	92.9	1179	3	US-09-176-666-46	Sequence 46, Appl
44	2131	92.9	1182	3	US-09-176-666-45	Sequence 45, Appl
45	2131	92.9	1188	3	US-09-176-666-44	Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-15
; Sequence 15, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-09-993-059-15

Alignment Scores:
Pred. No.: 2,19e-249
Score: 2293.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 3
Length: 1254
Matches: 417
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-15 (1-1254)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGGATGGCAGGACGCTT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTrpGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGGTATAGGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTCACAGCAAGGACTGAAGCTAGGAGATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACTTGGCAGGCTTCCCTGGAGATTTTGGATATCTACGACATTTGATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTGTGTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTTATAAGCACATGTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTCAAAGGCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTCATTCATTCCTGGAAAGATATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGAGCTGGACATCTTTTAAACGAGGAGAAATTTGTATGTTGCTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGACCCAGATATGTTAGTCAATTTGGCNACTTTGGCCTCAGCTGGATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTCATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuAsnLysAspValIleAlaIleAsn 320
DB 901 CGACACATACGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAAGGATACAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATATAACCCGCGAGGATTTGGT 1080
```

```
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGCTCCTCCCTGTGTAAAGGAAGCTAGGGTTCTTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTACATAATAATCCACAGGACTGTTTTCCTTCAGCTA 1251
```

RESULT 2

```
US-10-103-327-15
; Sequence 15, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-11-13
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-103-327-15
```

```
Alignment Scores:
Pred. No.: 2,19e-249 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
```

US-10-602-220-16 (1-417) x US-10-103-327-15 (1-1254)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGTTTCTTGGGACATCCCTGGGGCTAGAGCACTGCACAATCGANTTGGCAAGGACGCTT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACCTGCCAGGAAGGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAACTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAGGTATAGGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAAGAGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```



```
Db 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGGATTTATCAGATGTTGGAAATAAA 420
QY ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGAGATTGGATACACGATTTAGCCAGACTTTGCT 480
QY AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTGTGTACTGTGACAGTTTGGAAATTTG 540
QY AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACACATGCTCTGGCCCTGAATAGGACTGGCAGACATTTGTGTAC 600
QY SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCAGTGGCAATTTTGTCTGACATTCATGATTCCTGGAAAGATATAAG 720
QY SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGCTGGACATCTTTAAACCAGGAGAGAAATTGTGATGTGTGACCAGGG 780
QY GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db 781 GGTTGGNATGACCCAGATATGTTAGTATGTTGGCACTTTGGCTCAGCTGGAAATCAGCAA 840
QY ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
QY GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGACCCCTTGGGCAAGCAAGGTACCGTACGCTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAAACCGCAGGAGATTGGT 1080
QY GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
QY AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCCTCCCTGTGAAAGGAAGCTAGGGTCTATGAATGAGCT 1200
QY SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251
```

RESULT 3

```
US-09-993-059-11
; Sequence 11, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CUPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-09-993-059-11
```

Alignment Scores:

```
Prod. No.: 2,23e-249 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
```

US-10-602-220-16 (1-417) x US-09-993-059-11 (1-1266)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCCTCCCTGGCC 60
QY 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 120
QY 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrrLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrrMetAlaProGlnArg 100
Db 241 TGGAGGATGCGAGTTATGATGACTCTGTCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyValArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGAGAGACTTCAGGAGAGCCCTCAGCGCTTTCTCATGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCCTGGAGATTGGATACACGATTTAGCCAGACTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAenLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAAATTGATGGTGTGTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACACATGCTCTGGCCCTGAATAGGACTGGCAGACATTTGTGTAC 600
QY 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCAATCAGTGGCAATTTTGTCTGACATTCATGATTCCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGCTGGACATCTTTAAACCAGGAGAGAAATTGTGATGTGTGACCAGGG 780
QY 261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
```

```
Db      781  |||||GGTGGAAATGACCCAGATATGTAGTATGGCAACTTTGGCTCAGCTGGAATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerIleAsnLeu 300
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTTCATGTCTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuGlnAspIleAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGTAAATGGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGCAAGCAAGGTACCACTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGCTTAGCTGGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Qy      361  GlyProArgSerThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTGAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyTrpThr 400
Db      1141  GCCTGCTTCATCACAGCTCTCTCTCTGTTGAAAGGAGAGCTAGGTTCTATGATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201  TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 4

```
US-10-103-327-11
; Sequence 11, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; PRIOR FILING DATE: 2002-03-20
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1266
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1266)
US-10-103-327-11
```

Alignment Scores:

```
Pred. No.:      2,23e-249      Length:      1266
Score:          2293.00      Matches:      417
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              3              Gaps:      0
```

US-10-602-220-16 (1-417) x US-10-103-327-11 (1-1266)

```
Qy      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1  ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTCGCTTCCTGGCC 60
Qy      21  LeuValSerTrpAspIleProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
```

```
Db      61  CTCGTTTCTCTGGGACATCCTCTGGGGCTAGAGCACTGGACAATGGATTTGGCAAGGACGCT 120
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      121  ACCATGGGCTGGCTGCACCTGGGAGGCTTCATGTGCACACTTGCATGCCAGGAGAGCCA 180
Qy      61  AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      181  GATTCTTCATCAGTGAGAGAGCTTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGGC 240
Qy      81  TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      241  TGAAGAGATGCAAGTTATGAGTACCTTCGATTCATTCAGTGTGGATGGCTCCCAAGA 300
Qy      101  AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      301  GATTTCAGAAAGGACAGCTTCAGGACAGCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
Qy      121  AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTyTrpAlaAspValGlyAsnLys 140
Db      361  GCTAATTAATGTTACAGCAAGAGGACTGAAGCTAGGCAATTTATGCAGATGTGGAAATAAA 420
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160
Db      421  ACCTGGCAGGCTTCCCTGGGAGTTTGGATATACGACATTCATGCCAGACCTTTGCT 480
Qy      161  AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyTrpCysAspSerLeuGluAsnLeu 180
Db      481  GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGCTTTGGAAATTTG 540
Qy      181  AlaAspGlyTyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTy 200
Db      541  GCAGATGGTTATTAAGCACATGCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy      201  SerCysGluTrpProLeuTyTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db      601  TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
Qy      221  GlnTyCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      661  CAGTACTGCAATCACCTGGCGAAATTTGCTGACATTCATGATGATTCCTGGGAAAGTATTAAG 720
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      721  AGTATCTTGGACTGGACATCTTTTAAACCAGAGAGAAATTTGTGATGTGCTGACACAGGG 780
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      781  GGTTGGAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTATTCATGTCTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGATAAGGACGTAAATGGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCAGGAGAGATTGGT 1080
Qy      361  GlyProArgSerThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCTGTGAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyTrpThr 400
Db      1141  GCCTGCTTCATCACAGCTCTCTCTCTGTTGAAAGGAGAGCTAGGTTCTATGATGGACT 1200
```



```
; LOCATION: (1)...(1272)
US-10-103-327-17

Alignment Scores:
Pred. No.: 2,246-249 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-17 (1-1272)

Qy 1 MetGlnLeuAArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTCGCTTCCTGGGC 60

Qy 21 LeuValSerTrpAspLeuProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCATAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
Db 301 GATTTCAGAGGACGACTTCAGGAGAGAGGCTTCCTCATGGATTCGCCAGCTA 360

Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyLeuArgGlyLeuArgValGlyAsnLys 140
Db 361 GCTAATATGTTTACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTGGAAATAA 420

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpAspLeuAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATATGATGATGATGATGATGATGATGATGATG 480

Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTACATCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540

Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGGTTATAAGCACATGTCTTGGCCCTGAATAGGACTGGCAGAGCATGTGTGTAC 600

Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAGGCCAATATATACAGAAATCCGA 660

Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGCAATTTTGTGCTGACATGATGATCTTCTGGAAAAATATAAG 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTATGTTGTGGACACAGGG 780

Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGTGGATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCTCAGCTGGATCAGCAA 840

Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACATGATGGCCCTCTGGGCTCATATGGCTGCTCTTTATTCATGTCTAATGACCTC 900

Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaLeuAsn 320
Db 901 CGACATCATCAGCCCTCAGCCAAAGCTCTCTTCAGGTAAGAGCTAATTCCTCAAT 960

Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTACAGAGGAGAGCAACTTTGAAGTGTGG 1020

Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluLeuGly 360
Db 1021 GAACGACCTCTCTCAGGCTTACCTGGGCTGTAGCTATATATAACCGCAGGAGATTGGT 1080

Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTTCCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGAGTGGCCTGTAATCCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200

Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATNAATCCACAGGACCTGTTTGTCTCAGCTA 1251

RESULT 7
US-09-993-059-7
; Sequence 7, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-09-993-059-7

Alignment Scores:
Pred. No.: 2,266-249 Length: 1278
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-7 (1-1278)

Qy 1 MetGlnLeuAArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTCGCTTCCTGGGC 60

Qy 21 LeuValSerTrpAspLeuProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120

Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGCATCATAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
```

```
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCyseilAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTTGATGCAGATGTTGCCAGACTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGlnAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGTCCTGGCCCTGATAGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATTTGATGATTCCTGGAAAGATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACATGGACATCTTTTACCAGGAGAGAAATTTGTTGATGTTGCTGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACCAGCTTAGACAGGGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATAAACCGCGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGGAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCTT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGACTCTCCCTGTTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251
```

RESULT 8

US-10-103-327-7

; Sequence 7, Application US/10103327

; Patent No. 6890748

```
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1275)
US-10-103-327-7
```

Alignment Scores:

Pred. No.:	2,266-249	Length:	1278
Score:	2293.00	Matches:	417
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	3	Gaps:	0

US-10-602-220-16 (1-417) x US-10-103-327-7 (1-1278)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTCAGCAACCCAGCAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGGATGAGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATTATGTTTCACAGCAAGGAGCTGAAGCTAGGGAATTTATGCAGATGTTGGAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCTGGGAGTTTGGATACCTACGACATTTGATGCAGATGTTGCCAGACTTTGCT 480
QY 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGlnAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGTCCTGGCCCTGATAGACTGGCAGAGCATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
```


Db 1021 GAACGACCTCTCTAGGCTTAGGCTGGCTGTAGCTATGATAAACCGGACGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCATGATGAGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251

RESULT 10
US-10-103-327-13
; Sequence 13, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1284
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1284)
US-10-103-327-13

Alignment Scores:
Pred. No.: 2,27e-249 Length: 1284
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-10-103-327-13 (1-1284)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCTCTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGAACAATGATGGCAAGGACGCCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATCAGTGAGAGCTTCTCATGGAGTGGCAGAGCTCATGGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGTTTATGATACCTCTGCAATGATGACTGTGGATGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAAATTATGTTTACACAGCAAGGAGCTGAAGCTAGGATTTATGCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATATACAGACATTGATGCCAGACTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCGTGTAGTGGCTCTTTATATATGTGGCCCTTTCAAAGGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCATACACTGGCGAAATTTTGTGCACTGATGATTCCTGGAAAGATATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTGCTGGACCGAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGATACGACA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTTATCATGTCTATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATGGCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGCTACCACTTAGACAGGAGACAACCTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGGAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCATGATGAGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251

RESULT 11
US-09-993-059-3
; Sequence 3, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06

```
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1290)
US-09-993-059-3

Alignment Scores:
Pred. No.: 2,29e-249 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-3 (1-1290)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTCGCGCTTCCTGCGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCCTGGGACATCTCTGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCACTGTCAACCTTGACTGCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCAGTCAGAACTCTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGAAGGATGCAAGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
DB 301 GATTCAGAGCAGAGCTTCAGGCGAGCCCTCAGCGCTTCTCATGGATTGCCAGCTA 360

QY 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
DB 361 GCTAAATTATCTCACAGCAAGGACTGAAGCTAGGGATTTATGACATGTTGGAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCGAGCTTCCCTGGGAGTTTGGATATCTACGACATTTGATGCCAGAGCTTGTCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGTTTGTACTGTGACATTTGGAAAAATTG 540

QY 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGTTTATAAGCACATGCTCTGGCCCTTGAATAGGACTGGCAGAAAGCATTTGTATC 600

QY 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATATGCGCCCTTTTCAAAAGGCCCAATTATACAGAAATCCGA 660

QY 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCATCTGCGGAAATTTTGTGTGACATTCATGATTCCTCGAAAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 1 ATGCAGCTGAGGAACCCAGAACTTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGCGCC 60
```


Qy 21 LeuValSerTrpAspIleProGlyAlaAraAlaLeuAaspAenGlyLeuAlaAraGThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGATGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluAraGlyPheMetCysAenLeuAraGlyGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAaspCysTrpMetAlaProGlnAra 100
Db 241 TGGAAAGATGAGGTTATGATGATCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyAraGlnAlaAaspProGlnAraGlyPheProHisGlyIleAraGlnLeu 120
Db 301 GATTCAGAGCGAGACTTCAGGAGACCCCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAaspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGAGCTCAAGCTAGGAGTTTATGAGATGTTGGAAATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAaspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGCTTCCCTGGGAGTTTGGGATATACGACATGATGATGATGATGATGATGATGAT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAaspSerLeuGluAenLeu 180
Db 481 GATCGGGAGTAGATCTGCTAAATTTGATGGTGTGTACTGTGACATTTGGAAATTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAenAraGlyTrpGlyArgSerIleValTrp 200
Db 541 GCAGATGTTTATAGACACATGCTTGGCCCTGAATAGACTGGCAGACATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAenTrpThrGluIleAra 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAenHisTrpAraGlnPheAlaAaspIleAaspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGCAATGATGATGATGATGATGATGATGATGATGAT 720
Qy 241 SerIleLeuAaspTrpThrSerPheAenGlnAraGlyValAaspValAlaGlyProGly 260
Db 721 AGTATCTTGACCTGGACATCTTTTAAACCAGAGAGAAATGTTGATGTTGCTGACCCAGGG 780
Qy 261 GlyTrpAenAaspProAaspMetLeuValIleGlyAenPheGlyLeuSerTrpAenGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAaspProLeuPheMetSerAenAaspLeu 300
Db 841 GTAACCTCAGATGGCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAaspLysAaspValIleAlaIleAen 320
Db 901 CGACACATCAGCCCTCAGCCAAAGCTCTCTTCAGATTAAGACCGTAATTTGCCATCAAT 960
Qy 321 GlnAaspProLeuGlyLysGlnGlyTrpGlnLeuAraGlnGlyAaspAenPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTG 1020
Qy 341 GluAraProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAenAraGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATAAACCGCAGGAGATTTGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAenPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAAATCT 1140

Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysAraGlyLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAenProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAAATCCACAGGCACCTGTTTGTCTTACAGCTA 1251
RESULT 13
US-09-993-059-9
; Sequence 9, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1296
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1296)
US-09-993-059-9
Alignment Scores:
Pred. No.: 2,31e-249 Length: 1296
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0
US-10-602-220-16 (1-417) x US-09-993-059-9 (1-1296)
Qy 1 MetGlnLeuAraAenProGlnLeuHisLeuGlyCysAlaLeuAlaLeuAraGlyPheLeuAla 20
Db 1 ATGCAGCTCAGGAACCCAGAACCTACATCTGGGCTGGCGCTTGGCTTGGCTTGGCTTGGCTTGGCT 60
Qy 21 LeuValSerTrpAspIleProGlyAlaAraGlyValAaspAenGlyLeuAlaAraGThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGATGGCAAGGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluAraGlyPheMetCysAenLeuAraGlyGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAaspAlaGlyTrpGluTrpLeuCysIleAaspAaspCysTrpMetAlaProGlnAra 100
Db 241 TGGAAAGATGAGGTTATGATGATCTCTGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyAraGlnAlaAaspProGlnAraGlyPheProHisGlyIleAraGlnLeu 120
Db 301 GATTCAGAGCGAGACTTCAGGAGACCCCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAaspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGAGCTCAAGCTAGGAGTTTATGAGATGTTGGAAATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAaspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGCTTCCCTGGGAGTTTGGGATATACGACATGATGATGATGATGATGATGATGAT 480

		; ORGANISM: Homo sapiens	
		; FEATURE:	
		; NAME/KEY: CDS	
		; LOCATION: (1) ... (1296)	
		US-10-103-327-9	
		Alignment Scores:	
		Pred. No.:	2,31e-249
		Score:	2293.00
		Length:	1296
		Matches:	417
		Conservative:	0
		Mismatches:	0
		Best Local Similarity:	100.00%
		Indels:	0
		Query Match:	100.00%
		Gaps:	0
		DB:	3
		US-10-602-220-16 (1-417) x US-10-103-327-9 (1-1296)	
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
DB	1	ATGCAGCTGAGGAAACCAGAACTACATCTGGGCTGGCGCTTGGCGCTTCTCTTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
DB	61	CTCGTTCTCTGGGACATCCCTGGGCTAGAGCACGGACAATGGATTGGCAAGAGCCCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
DB	121	ACCATGGGCTGGCTGCATCTGGAGGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
DB	181	GATTCTCGCATCAGTGAGAAAGCTTTCATGGAGATGGCAGAGCTCATGTCTCAGAAGGC	240
QY	81	TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg	100
DB	241	TGCAAGGATGCAAGTTATGAGTACCTTCGATTCATGATGACTGTGTGGATGGCTCCCCAAGA	300
QY	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
DB	301	GATTTCAGAGGCGAGACTTCAGGCGAGCCCTCAGGGCTTTCTCATGGGATGGCCAGCTA	360
QY	121	AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys	140
DB	361	GCTAATTTATGTTTACAGCAAGAGCTGAAAGCTAGGAGTTTATGTCAGATGTGGAAATAAA	420
QY	141	ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla	160
DB	421	ACCTGGCGAGGCTTCCTCTGGGAGTTTGGATATCTACGACATTCGTCAGATTCGCCAGACTT	480
QY	161	AspTrpGlyValAspLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu	180
DB	481	GACTGGGAGTAGATCTGCTTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTGT	540
QY	181	AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp	200
DB	541	GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATTGTGTAC	600
QY	201	SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg	220
DB	601	TCCTGTGAGTGGCTCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATCCGA	660
QY	221	GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys	240
DB	661	CAGTACTGCAATCACTGGCGGAAATTTTGTGACATTTGATGATTTCTCTGGAAAAGTATAAG	720
QY	241	SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
DB	721	AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTGTGTGGACCAAGG	780
QY	261	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln	280
DB	781	GGTTGGAAATGACCCAGATATGTAGTATGGCACTTTGGCCCTCAGCTGGATCAGCAA	840
QY	281	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
DB	841	GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTATGATCTAATGACCTC	900
QY	301	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
DB	901	CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCCATCAAT	960
QY	321	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
DB	961	CAGGACCCCTTGGGCAAGCAGGATACCGCTTAGACAGGAGACAACTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
DB	1021	GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT	1080
QY	361	GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
DB	1081	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAACTCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr	400
DB	1141	GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAAGGAGCTAGGGTTCTATGATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
DB	1201	TCAAGGTTAAGAGTCACATAAATCCACAGGCAGCTGTTTGTCTTCAGCTA	1251
		RESULT 14	
		US-10-103-327-9	
		; Sequence 9, Application US/10103327	
		; Patent No. 6890748	
		; GENERAL INFORMATION:	
		; APPLICANT: GARGER, Stephen A.	
		; APPLICANT: TURPEN, Thomas H.	
		; APPLICANT: KUMAGAI, Monto H.	
		; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN	
		; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION	
		; FILE REFERENCE: 008010087CPUS06	
		; CURRENT APPLICATION NUMBER: US/10/103,327	
		; CURRENT FILING DATE: 2002-03-20	
		; PRIOR APPLICATION NUMBER: US/09/993,059	
		; PRIOR FILING DATE: 2001-11-13	
		; NUMBER OF SEQ ID NOS: 37	
		; SOFTWARE: FastSeq for Windows Version 4.0	
		; SEQ ID NO 9	
		; LENGTH: 1296	
		; TYPE: DNA	

Db 841 GTAACATGAGATGGCCCTCTGGGCTATCATGGCTGCTCTTATTTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAGGCCAAAGCTCTCTTCAGGATTAAGACGTATTTGGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTTP 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTAGACAGGAGCAACCTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTTPAlaValAlaMetIleAsnArgGlnGlnGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGTGGCTGTAGCTATGATAAACCGGCAGGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTTPThr 400
Db 1141 GCCTGTCTCATCACAGCTCCCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 15

US-09-993-059-5
; Sequence 5, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1308)
US-09-993-059-5

Alignment Scores:
Pred. No.: 2,34e-249 Length: 1308
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-602-220-16 (1-417) x US-09-993-059-5 (1-1308)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTGCGCTTCTCGCTTCTCGGCC 60
Qy 21 LeuValSerTTPAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCGGACATCCCTCGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
Qy 41 ThrMetGlyTTPLeuHisTTPGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGACCTGGGAGCGCTTCATGTGCACCTTGTGACTGCCAGGAAGAGCCA 180

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TTPLeuAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTTPMetAlaProGlnArg 100
Db 241 TGGAAAGATGCGAGTTATGAGTACCTCTGCAATTGATGACTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGCAGACTTCAGGCAGACCTCAGCGCTTCTTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTTTCACAGCAAGGACTGAAAGCTAGGATTTATGACAGATGTGGAAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACACGACATTCATGCCAGACTTGGCT 480
Qy 161 AspTTPGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
Qy 201 SerCysGluTTPProLeuTyrMetTTPProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTCACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTTPArgAsnPheAlaAspIleAspAspSerTTPLysSerIleLys 240
Db 661 CAGTACTGCAATCAGCTGGCGAATTTTGTGTCATGTATGATTCCTGGAAAGATATAAG 720
Qy 241 SerIleLeuAspTTPThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACACAGGAGAGAAATTTGATGTCTGCTGACACAGG 780
Qy 261 GlyTTPAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTTPAsnGlnGln 280
Db 781 GGTGGATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTTPAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAGGCCAAAGCTCTCTTCAGGATAAGGACGTAAATTCGCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTTP 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTACCAAGCTTAGACAGGAGCAACCTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTTPAlaValAlaMetIleAsnArgGlnGlnGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGTGGCTGTAGCTATGATAAACCGGCAGGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTTPThr 400
Db 1141 GCCTGTCTCATCACAGCTCCCTCCCTGTGAAAAGGAAGCTAGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

Tue Jan 3 11:36:07 2006

Search completed: January 1, 2006, 05:53:05
Job time : 367.352 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:17 ; Search time 7564.48 Seconds
(without alignments)
2579.189 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MQLRNPELHGICALALRFLA.....EWTSLRSHINPTGTLLQL 417

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q/cgn2.1/USPTO_spool/US1062220/runat 30122005 140558 5433/app_query.fasta_1.782
-DB=EST -QFMT=fascap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1062220 @CGN 1.1 6731 @runat 30122005 140558 5433 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

```
EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1266	4 CR607242	CR607242 full-length
2	2293	100.0	1277	4 CR605654	CR605654 full-length
3	2289	99.8	1253	4 CR617861	CR617861 full-length
4	2083	90.8	1290	10 AY408540	AY408540 Homo sapi
5	2034	88.7	1290	10 AY408541	AY408541 Pan trogl
6	1806	78.8	1376	4 AK054547	AK054547 Mus muscu
7	1806	78.8	2962	4 AK040081	AK040081 Mus muscu

8	1777	77.5	1086	1	AL554978	AL554978
9	1699	74.1	1014	5	EX354096	EX354096
10	1691.5	73.8	999	1	AL552630	AL552630
11	1667.5	72.7	1067	3	BM564282	BM564282
12	1659.5	72.4	1005	3	BQ062192	AGENCOURT
13	1649	71.9	922	7	CO645623	ILLUMIGEN
14	1648.5	71.9	1026	1	AL577581	AL577581
15	1627	71.0	1296	10	AY408542	AY408542 Mus muscu
16	1620.5	70.7	1133	7	CO645672	ILLUMIGEN
17	1589	69.3	920	7	CO645464	ILLUMIGEN
18	1578	68.8	923	7	CT005156	CT005156
19	1574	68.5	984	7	CO646251	ILLUMIGEN
20	1556.5	67.9	976	5	BQ956043	AGENCOURT
21	1527	66.6	898	5	BU191867	AGENCOURT
22	1519	66.2	922	5	EU540848	AGENCOURT
23	1515.5	66.1	937	5	BQ934640	AGENCOURT
24	1492.5	65.1	852	6	CA454083	AGENCOURT
25	1465	63.9	888	5	BU154569	AGENCOURT
26	1458	63.6	958	6	CA487415	AGENCOURT
27	1452	63.3	871	6	CA454143	AGENCOURT
28	1443	62.9	801	5	BU596617	AGENCOURT
29	1429	62.3	1025	3	BM450649	AGENCOURT
30	1425.5	62.2	849	6	CA487531	AGENCOURT
31	1422.5	62.0	879	2	BG824387	602728627
32	1412.5	61.6	927	2	BE622583	601440703
33	1408.5	61.4	871	3	BQ225444	AGENCOURT
34	1406	61.3	908	2	BE379425	601237275
35	1404	61.2	796	8	DR156332	HESC2_64
36	1402.5	61.2	921	6	CA488935	AGENCOURT
37	1397	60.9	845	7	CR983967	CR983967
38	1395.5	60.9	1071	1	AL575861	AL575861
39	1384	60.4	813	5	EX344841	EX344841
40	1382	60.3	782	2	B1224248	602940538
41	1371.5	59.8	1066	2	BG824323	602727346
42	1365	59.5	749	5	EX344842	EX344842
43	1362	59.3	908	6	CD385342	AGENCOURT
44	1360	59.3	886	3	BP435628	BP435628
45	1303	56.8	767	3	B1753664	603023559

ALIGNMENTS

RESULT 1

CR607242

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CR607242 1266 bp mRNA linear HTC 21-JUL-2004
full-length cDNA clone CSODI067VJ01 of Placentia Cot 25-normalized
of Homo sapiens (human).

CR607242 GI:50488049

HTC; CNSLT_CDNA.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1266)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished

Contact: Feng Liang Email: fliang@lifetech.com URL:

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue

2 (bases 1 to 1266)

Genoscope.

Direct Submission

Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage:

BP 191 91006 Evry cedex - FRANCE (E-mail: seque@genoscope.cns.fr

- Web: www.genoscope.cns.fr)

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen.

```
FEATURES
source
Location/Qualifiers
1..1266
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1067YJ01"
/tissue_type="Placenta Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
Pred. No.: 1,21e-240 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-16 (1-417) x CR607242 (1-1266)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 7 ATGCAGCTGAGAAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCGCTTCCTGGCC 66
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 67 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGATGGCAGGACGCT 126
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 127 ACCATGGCTGGCTGCATGGGAGCGCTTCATGTGCAACTTGACTGCCAGGAGAGCCA 186
Qy 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 187 GATTCTCTGCATCAGTGAAGAGCTTTCATGAGATGGCAGAGCTCATGGTCTCAGAGGC 246
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 247 TGGAGAGATGACAGTTATGAGTACTCTGCATTTGATGATGATGATGGTCCGCCAAGA 306
Qy 101 AspSerGluArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 307 GATTCAAGAGCAGACTTCAGGACAGCCCTCAGCGCTTCTCATGGGATTCGCCAGCTA 366
Qy 121 AlaSerTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 367 GCTAAATTATGTTACAGCAAGAGAGTGAAGTGGGATTTATGAGATGTTGGAAATATA 426
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 427 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTCATGGCAGCCTTTGCT 486
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 487 GACTGGGAGTAGATCTGCTAAATTTGATGGTTTACTGTGACAGCTTTGGAAATTTG 546
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 547 GCAGATGTTTATAAGCAGCATGCTCTGGCCCTGATAGGACTGGCAGAGCATTTGTGAC 606
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpTrpGluIleArg 220
Db 607 TCCTGTAGTGGCTCTTTTATATGTGGCCCTTTCAAAGGCCAATTTATACGAATTCGA 666
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 667 CAGTACTGCAATCACTGGCGAAATTTTGTCTGACATTCATGATGATTCCTCGAAAGTATAAG 726
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 727 AGTATCTTGGACTCGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTGGACAGGG 786
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280

FEATURES
source
1..1277
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D808YB11"
/tissue_type="Neuroblastoma Cot 10-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Alignment Scores:
```

```
Pred. No.: 1,22e-240 Length: 1277
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-602-220-16 (1-417) x CR605654 (1-1277)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 10 ATGCAGCTGAGGACCCAGACACTACATCTGGCTGGCGCTTGGCTTCTTCTGGCC 69
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArThrPro 40
Db 70 CTGCTTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCT 129
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 130 ACCATGGGCTGGCTGCACTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 189
Qy 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 190 GATTCCTGCATCAGTGAAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGC 249
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 250 TGGGAAGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 310 GATTCAGAAGGACACTTCAGGAGACCTCAGGCGCTTTCCTCATGGGATTTCGCCAGCTA 369
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 370 GCTAATATGTTACAGCAAGGACTGAGTAGGATTTATGAGATGTTGGAAATATA 429
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
Db 430 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACGACATGATGATGATGATGATGATGAT 489
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 490 GACTGGGAGTAGATCTGCTAATAATTTGATGGTTTGTACTGTGACGATTTGGAAATTTG 549
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 550 GCAGATGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC 609
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTrpThrGluLeuArg 220
Db 610 TCCTGTGAGTGGCTCTTTATATGTGGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 669
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 670 CAGTACTGCAATCACTGGCGAAATTTTGTGCTACATGATGATGATGATGATGATGATGAT 729
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 730 AGTATCTGGACTGGACATCTTTTAAACGAGAGAGAAATTTGTATGTTGTCTGGACAGGG 789
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 790 GGTGTGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 849
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 850 GTAACTCAGATGGGCTCTGGGCTATCATGCTGCTCTCTTTTATTCATGCTCTTAATGACCTC 909
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 910 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAGGACGTAATTTGCCATCAAT 969
```

```
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 970 CAGGACCCCTTGGGCAAGGAGGTACAGCTTAGACAGGAGAGCAACTTTTGAAGTGTGG 1029
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1030 GNACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCCGCGAGGATGGT 1089
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1090 GGACCTCGCTTATACCATCGCAGCTTCTCCCTGGTAAAGGAGTGGCTGTAAATCCT 1149
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1150 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTAGGTTCTATGAATGGACT 1209
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1210 TCAAGTTTAGAGTCAATATATCCACAGGCACTGTTTGTCTTCAGCTA 1260

RESULT 3
CR617861 1253 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CSODK009YE12 of HeLa cells Cot 25-normalized
of Homo sapiens (human).
ACCESSION CR617861
VERSION CR617861.1 GI:50498668
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1253)
AUTHORS Li, W.B., Gruber, C., Jessup, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact: Feng Liang Email: fliang@lifetech.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paraday Avenue
REFERENCE 2 (bases 1 to 1253)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life technologies, a
division of Invitrogen.
FEATURES
source
1..1253
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODK009YE12"
/tissue types="HeLa cells Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Alignment Scores: 3,27e-240 Length: 1253
Pred. No.: 2289.00 Matches: 416
Score: 2289.00 Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 99.83% Gaps: 4
DB: 4
US-10-602-220-16 (1-417) x CR617861 (1-1253)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
```

```
Db      6  ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTGCGCTTCTCTGGCC 65
Qy      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      66  CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCCT 125
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      126  ACCATGGCTGGCTGGCACTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGACCA 185
Qy      61  AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      186  GATTCTCGATCAGTGAGAGCTCTTCATGAGATGCGAGAGCTCATGGTCTCAGAAGGC 245
Qy      81  TrpIysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      246  TGGAAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 305
Qy      101  AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db      306  GATTTCAGAAGCGACATTCAGCGAGACCTCAGCGCTTTCTCATGGATTGCGACGTA 365
Qy      121  AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db      366  GCTAATTATGTTTCACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA 425
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db      426  ACCTGCGCAGGCTTCCCTGGGAGTTTTGGTACTACGACATTTGATGCCAGACCTTTGCT 485
Qy      161  AspTrpGlyValAspLeuLeuPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db      486  GACTGGGAGTAGATCTGCTAATAATTTGATGGTTGTACTGTGACAGTTGGAAAAATTG 545
Qy      181  AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db      546  GCAGATGTTTATAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTAC 605
Qy      201  SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db      606  TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTATACAGAAATCCGA 665
Qy      221  GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      666  CAGTACTGCATCACTAGCGGAAATTTTCTGACATTTGATGATCTCTGGAAAGATTAAG 725
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      726  AGTATCTTGGACTGGACATCTTTTAAACCAGCAGAGAAATTGTTGATGTTGCTGGACAGG 785
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      786  GGTTCGAATGACCCAGATATGTTAGTGAATGGCAACTTTTGGCCTCAGCTGGAAATCAGCA 845
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      846  GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTCTTTATCATGCTTAATGACCTC 905
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      906  CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT 965
Qy      321  GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db      966  CAGGACCCCTTTGGCAAGCAAGGGTACCAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1025
Qy      341  GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnIleGly 360
Db      1026  GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATTAACCCGCGCAGAGATTGGT 1085
Qy      361  GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1086  GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCCTGTAATCCT 1145
```

```
Qy      381  AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTrpGluTrpThr 400
Db      1146  GCCTGCTTCATCATCACAGCTCCCTCCCTGTGAAAAGGAGTAGGTTCTATGATGGACT 1205
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGln 416
Db      1206  TCAAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTTGCTTCAG 1253

RESULT 4
AY408540      1290 bp      DNA      linear      GSS 15-DEC-2003
LOCUS      Homo sapiens GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION      genomic survey sequence.
ACCESSION      AY408540
VERSION      AY408540.1 GI:39764511
SOURCE      GSS.
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE      1 (bases 1 to 1290)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
              Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
              Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
              Adams,M.D. and Cargill,M.
              Inferring nonneutral evolution from human-chimp-mouse orthologous
              gene trios
              Science 302 (5652), 1960-1963 (2003)
JOURNAL      14671302
PUBMED      2 (bases 1 to 1290)
AUTHORS      Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
              Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
              Perriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
              Adams,M.D. and Cargill,M.
              Direct Submission
              Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
              Rockville, MD 20850, USA
              This sequence was made by sequencing genomic exons and ordering
              them based on alignment.
FEATURES      Location/Qualifiers
source      1..1290
              /organism="Homo sapiens"
              /mol_type="genomic DNA"
              /db_xref="taxon:9606"
gene      <1..>1290
              /gene="GLA"
              /locus_tag="HCM3258"
ORIGIN
Alignment Scores:      1.51e-217      Length:      1290
Pred. No.:      2083.00      Matches:      386
Score:
Percent Similarity:      92.57%      Conservative:      0
Best Local Similarity:      92.57%      Mismatches:      31
Query Match:      90.84%      Indels:      0
DB:      Gaps:      0

US-10-602-220-16 (1-417) x AY408540 (1-1290)
Qy      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1  ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCGCTTCTCTGGCC 60
Qy      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      61  CTCGTTTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGGATTGGCAAGACGCCT 120
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db      121  ACCATGGCTGGCTGGCACTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGACCA 180
```

```
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGAGATGAGGTTATGAGTACTCTGCAATGATGACTGTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTGAGAAGGCAGCTTCAGGCAGACCTTCAGCGCTTCTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaSerTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTCACAGCAAGAGCTGAAGCTAGGATGTTATGACAGATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATGACTACGACATTTGATGGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGGTAGATCTGCTAAATTTGATGTTGTTACTGTGACAGTTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATTCATGATCTCTGCGAAAGATATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTGGACTGGACATCTTTTAAACAGCAGAGAGAATTTGTGATGTTGCTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTTGGAAATGACCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTATTCATGCTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAAGCAGGTTACCGCTTAGACAGGGAGACAACTTTGAAGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATATAAACCGGAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaLysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTGAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTTCATCACACAGCTCCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACTAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 5

```
AY408541 1290 bp DNA linear GSS 15-DEC-2003
LOCUS Pan troglodytes GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408541
VERSION AY408541.1 GI:39764512
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Pan.
REFERENCE 1 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1290)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferrera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..1290
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1290
/gene="GLA"
/locus_tag="HCM3258"
gene
ORIGIN
Alignment Scores: 3.66e-212 Length: 1290
Pred. No.: 2034.00 Matches: 377
Score: 2034.00
Percent Similarity: 90.89% Conservative: 2
Best Local Similarity: 90.41% Mismatches: 38
Query Match: 88.70% Indels: 0
DB: 10 Gaps: 0
US-10-602-220-16 (1-417) x AY408541 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGCGGCTTGGCTTCCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCTTTTCTGGGACATCCCTTCGGGCTAGACACTGAGCAATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGATCTGCCAGGAAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAAGCTCTTCATGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGAGATGAGGTTATGAGTACTCTGCAATGATGACTGTGTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
```



```
Db      301 GATTTCAGAGCGAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATCGGATTCGNNAGCTA 360
Qy      121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db      361 NNNNNNNNGTTTCAGCAAGAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db      421 ACCTGCCGAGCTTCCCTGGGAGTTTTGGTACTACGACATGTATGCCAGACCTTTGCT 480
Qy      161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db      481 GACTGGGAGTAGATCTGCTAAATTTTGGTGTGTTACTGTGACAGCTTTGGAAATTTG 540
Qy      181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db      541 GCAGATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy      201 SerCysGluTyrProLeuTyrMetTyrPheGlnLysProAsnTyrThrGluIleArg 220
Db      601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy      221 GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys 240
Db      661 CAGTACTGCAATCACTCGCGAAATTTTGTGCTGACGTTGATGATTCCTGCGAAAGTATAAG 720
Qy      241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      721 AGTATCTTGGACTGGACATCTTTTAAACGAGAGAGAAATTGTGTATGTTGCTGGACCAAGG 780
Qy      261 GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln 280
Db      781 GGTGGAGTACCCAGATATGTTAGTATGTTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy      281 ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTCTCTTTTATTATTCATGCTAATGACCTC 900
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy      321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      961 CAGACCCCTTGGCAAGCAGAGGTACAGCTTAGACAGNAGCAACTTTGAAGTGTGG 1020
Qy      341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCAGGAGATTTGGT 1080
Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081 GGACCTCGCTCTTATACCATCGCACTTGCTTCCCTGGGTAAAGAGTGGCGCTGTAATCCT 1140
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      1141 GCCTGTCTTCATCACACAGCTCCTCCCTGTGTAAGGAGCTAGGCTTCTATGAATGACT 1200
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201 TCNAGGTTAAGAGTCAATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 6
AK054547
LOCUS
DEFINITION Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
full-length enriched library, clone:E330039P08
product:galactosidase, alpha, full insert sequence.
ACCESSION AK054547
VERSION AK054547.1 GI:26344324
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
```

```
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE
AUTHORS Carninci,P., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
PUBMED 11076861
REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayaehida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
source
1..1376
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6J"
/db_xref="FANTOM_DB:E330039P08"
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.


```

/db_xref="taxon:10090"
/clone="E330039P08"
/sex="female"
/tissue_type="ovary"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="2 days pregnant adult"
21..1286
/note="unnamed protein product; galactosidase, alpha
(MGI:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC35819.1"
/db_xref="GI:26344325"
/translaton="MAMKLSRDLVCEALPLALFWSILGVRALDGLARTPTM
GWLHWFMCNLDQCEPFDACISQELFMQMAELMVDGWRDAGDYDILCIDDCWAPR
DSKRLQADPORPSGKHLANYVHSKGLGIVADVGNKTCAGFGSGYDIDAOT
PADWGDLKPEDGCHCDVSVLENGYKYMALANRGTGRSVYSCWENPLYLRPHKPY
TDQYCNHWRNFDVDVDSWESIKNLSWTVVYQKEIVVAGPQSNPDLNIGNFG
LSWDQYQVQWALMAIMAPLMSNDLRQISQAKALLQNDVIAINODPLGKQGYCFR
KENHIEYWERPLSNLAWAVRNLOETGGPCPTIOTISLGRGLACNPGCIIITQLLPE
KVHLGFVEWTLTLKTRVNSPGTVLFLRER"
1358..1363
/note="putative"
1376
/note="putative"

```

CDS

polyA_signal

polyA_site

ORIGIN

```

Alignment Scores:
Pred. No.: 4,67e-187 Length: 1376
Score: 1806.00 Matches: 326
Percent Similarity: 88.01% Conservative: 41
Best Local Similarity: 78.18% Mismatches: 50
Query Match: 78.76% Indels: 0
DB: 4 Gaps: 0

```

US-10-602-220-16 (1-417) x AK054547 (1-1376)

```

Qy 1 MetGlnLeuArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 27 ATGAAGCTTTTGACGAGATACCGGCTGTCTGTGAGCTTGGCTTTGCTCCCTGGCT 86
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAnglyLeuAlaArgThrPro 40
Db 87 TTAGTCTTCTGGAGCAATCTTGGGCTCAGACATTTGGCAATGCTTTGGCGGAGCTCCT 146
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysLeuLeuAspCysGlnGluPro 60
Db 147 ACTATGGGCTGGCTGCAATTTGGAAACGTTTATGTGCAACCTTTGACTGCCAAGAGAGCCT 206
Qy 61 AspSerCysIleSerGluLeuLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 207 GATGCTGCATAGTGAGCACTGTTTCAGATGAGGAGCTCATGCTCTGTATGGC 266
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 267 TGGCGGATGCAGGTTATGATCTCTGATAGTACGCTTGTGATGCTTCCGAGAGG 326
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 327 GATTCAAGGCGCAGGCTTACGGCAGATCCCAACGCTTTCTAGTGGGATCAACACCTC 386
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 387 GCAATTTAGTCCACACCAAGGATTTAGCTAGGATTTATGCGAGATTTGGGATATAA 446
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 447 ACCTGTGCAGGTTTCCCGGAGTTTGGATCTTATGACATTTGATGCGCAGACATTTGCT 506
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 507 GACTGGGCGGTAGATCTGCTAAATTTGATGGTTGTCACTGTGACAGTGTAGTATCTCTTG 566

```

```

Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 567 GAGAAATGTTATAAGTACATGGCCCTTGAACAGGACAGCCGCAAGCATTTGTATAC 626
Qy 201 SerCysGlnTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpTrpGluLeuArg 220
Db 627 TCCTGTGAGTGGCCACTTTATTTAGAGCCCTTTTCATAGGCCCAATATATACAGATATCAA 686
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
Db 687 TATTACTGCAATCATTTGGAGAAATTTTATGATGTTTATGATCTCTGGAAAGCATAAAG 746
Qy 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 747 AATATCTTGTCTTGGACAGTGGTTTACAGAAAGAGATTTGTTGAAGTCGCTGGACCAAGC 806
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 807 AGCTGGAAATGACCCAGACATTTAGTGTATGCGCAACTTTTGGCCTCAGTTGGAGCAGCAG 866
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 867 GTGACACAGATGGCCCTCTGGGCTATCATGGCCCTCCCTACTCATGTCCCAACGATCTG 926
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 927 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTTCAGAAATAAGGATGTAATTCCTCAAC 986
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 987 CAAGACCCCTTGGGCAAGCAGGCTACTGTTTCAGAAAGGAAACCAACCATTTGAGTTGG 1046
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1047 GAACGGCCACTCTCCAACCTAGCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATTGGT 1106
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1107 GGACCTTGTCTTATACCATCCAGATTTCTTCCCTGGGTAGAGGACTAGCTGCAATCCT 1166
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1167 GGTGTGATCATTTACTAGCTTCTCCCGAGAAAGTACACATGAGCTTCTATGAATGAGCT 1226
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1227 TTGACCTTAAAACTCGATTAAACCTCTCAGGCACTGTTTGTTCGGTTA 1277

```

RESULT 7

```

AK040081 2962 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DEFINITION library, clone:A430057F16 product:galactosidase, alpha, full insert
sequence.
ACCESSION AK040081
VERSION AK040081.1 GI:263333580
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes

```

```

JOURNAL
PUBMED
REFERENCE
AUTHORS
Genome Res. 10 (10), 1617-1630 (2000)
1042159
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kiseunai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2962)
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M.,
Hayaashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohnato,N.,
Okazaki,Y., Saito,K., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahiira,S.,
Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
Muramatsu,M. and Hayaashizaki,Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayaashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.jp/
URL:http://fantom.gsc.riken.jp/.
FEATURES
source
Location/Qualifiers
1..2962
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM,DB:A430057F16"
/db_xref="taxon:10090"
/clone="A430057F16"
/tissue_type="thymus"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1..1266
/note="unnamed protein product; galactosidase, alpha
(MGD|MG1:1347344, GB|NM_013463, evidence: BLASTN, 99%,
match=1373)
putative"
/codon_start=1
/protein_id="BAC30508.1"
/db_xref="GI:26333581"
CDS
/translation="MAWKLLSRDRLVCEALCLALVFWISLIGVRLDNGLARTPTM
GTLWRFMCNLDCEPEPACISEQLFMQMAELMVSVDGWRDAGDYLCIDDCWAPER
DSKRLQADPQRPSPGIEKILANYHSGKLKGLGIYADVGNKTCAGFPSCFSGSYDIAOT
FADMGVDLLKFDGCHDVSVLENGKYMALALNRGTSIVYSCPEPLYLRPHKPNY
TDIQYCNHRNFDVDSWESIKNLSWTYYQKEIVVAGSVGNDPMLVIGNFG
LSWDQQYQNALWAIWAAPALLMSNDLRQISSQAKALLQNDVIAINQDPLCKQGYCFR
KENHIEVWRPISNLAWAVARNLQIGGPGPYTIQISSIGRLGLACNPNCIIITOLLEPE
KVHLGFYEWTLTLKTRVNPSTGLFLRLR"
ORIGIN
Alignment Scores:
Pred. No.: 1.5e-186 Length: 2962
Score: 1806.00 Matches: 326
Percent Similarity: 88.01% Conservative: 41
Best Local Similarity: 78.18% Mismatches: 50
Query Match: 78.76% Indels: 0
DB: 4 Gaps: 0
US-10-602-220-16 (1-417) x AK040081 (1-2962)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 7 ATGAAGCTTTTGGAGAGATACGGCGCTGCTGTGAGCTTGGCGCTTTGCTCCCTGGCT 66
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 67 TTAGTTTCTGGAGCAATCTTGGGGTCAGAGCATGGACAATGGCTTGGCGCGGACTCCT 126
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 127 ACTATGGCGCTGGCTGCAATGGGAACGTTTCATGTGCAACCTTGACTGCCAAGAGACCT 186
QY 61 AspSerCysIleSerGluLysLeuPheMetGluAlaGluLeuMetValSerGluGly 80
Db 187 GATGCCCTGCATAAGTAGAGCAACTGTTTCATGCAGATGGCAGAGCTCATGGTCTCGATGGC 246
QY 81 TrpLysAspAlaGlyTyTrpGluTyTrpLeuGlyLeuGlyValAlaAspValGlyAsnLys 100
Db 247 TGGCGGGATGCAGGTTATGACTATCTGTCATAGATGACTGTTGGATGGCTCCCGAGAGG 306
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 307 GATTCAAGGGCGAGCTTCAGGCAGATGCCCAACGCTTCTCTAGTGGGATCAACACCTC 366
QY 121 AlaAsnTyTrpValHisSerLysGlyLeuLysLeuGlyIleTyTrpAlaAspValGlyAsnLys 140
Db 367 GCAATTTACTCCACCAAGGATTCAGACTAGGAGTATTCGAGATGTTGGGAATAA 426
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyTrpAspIleAspAlaGlnThrPheAla 160
Db 427 ACCTGTGCAGGTTTCCCGGGAGTTTGGATCCTATGACATTCATGCGCAGACATTTGCT 486
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyTrpCysAspSerLeuGluAsnLeu 180
Db 487 GACTGGGGCGTAGATCTGCTAAAAATTTGATGGTGTGCACCTGTGACAGTGTAGTATCCTTG 546
QY 181 AlaAspGlyTyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyTr 200
Db 547 GAGAAATGTTTATAGTACATAGCGCTTGGCCCTTGACAGGACAGCCCGAGCATTTGTATAC 606
QY 201 SerCysGluTrpProLeuTyTrpMetTrpProPheGlnLysProAsnTyTrpGluIleArg 220
Db 607 TCCTGTGAGTGGCCACTTTTATTTGAGACCCCTTTTCATTAAGCCCAATTTATACAGATATCAA 666
QY 221 GlnTyTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 667 TATTACTGCAATCATTTGGAGAAATTTTGTGATGATGTTTATTCCTCGGAAAGCATTAAG 726
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGlyIleValAspValAlaGlyProGly 260
Db 727 AATATCTTGTCTTGGACAGTGGTTTACCAGAGAGAGATTTGTAAGTTCGCTGGACACGAC 786
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280

```

```

Db      787 AGCTGGAAATGACCCAGACATGTTAGTGATCGCAACTTTGGCCCTCAGTTGGGACACAGAG 846
Qy      281 ValThrGlnMetAlaLeuTrrAlaIleMetAlaAlaProLeuPheMetSerAspAspLeu 300
Db      847 GTGACACAGATGGGCCCTCTGGGCTATCATGGCCGCTCCCTACTCATGTCCAACGATCTG 906
Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      907 CGACAAATCAGCTCTCAAGCCAAAGCTCTGCTCAGAATAAGGATGTAATTTGCCATCAAC 966
Qy      321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrr 340
Db      967 CAAGACCCCTGGGCAACAGCGGTACTGTTTCAGAAAGGAAACCAACCATTTGAGTTGG 1026
Qy      341 GluArgProLeuSerGlyLeuAlaTrrAlaValAlaMetIleAsnArgGlnGlnIleGly 360
Db      1027 GAAGGGCACTCTCAACCTAGCTGGGCTGTGGCTGTGAGAAACCTGCAGGAGATTGGT 1086
Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1087 GGACCTTGCTTATACCATCCAGATTCTTCCCTGGGTAGAGGACTAGGCTGCAATCCT 1146
Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysGlyPheTyrGluTrrThr 400
Db      1147 GGCTGTCATCACTACTCAGCTTCTCCCGAGAAAGTACACCTAGGCTTCTATGAATGGACT 1206
Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1207 TTGACCTTAAAACTCGATNAACCCCTCAGGCACTGTTTGTTCGGTTA 1257

```

RESULT 8

AL554978

LOCUS

DEFINITION AL554978 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK009YE12 5-PRIME, mRNA sequence.

ACCESSION

AL554978

VERSION

AL554978.3

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1086)

Li, W.B., Gruber, C., Jeesse, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 15, 2001 this sequence version replaced gi:31276787.

Contact: Genoscope

Genoscope - Centre National de Sequencage

2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime

end enriched, double-strand cDNA was digested with Not I and cloned

into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library

was normalized. Library was constructed by Life Technologies, a

division of Invitrogen. This sequence belongs to sequence cluster

10506.r

For more information about this cluster, see

http://www.genoscope.cns.fr/cdna?8=CS0DK009BC06QPl&c=10506.r.

FEATURES

Location/Qualifiers

1..1086

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DK009YE12"

/cell_type="HELA CELLS COT 25-NORMALIZED"

/cell_line="HELA"

/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and EcoR V

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.: 5,02e-184 Length: 1086

Score: 1777.00 Matches: 342

Percent Similarity: 96.65% Conservativeness: 4

Best Local Similarity: 95.53% Mismatches: 10

Query Match: 77.50% Indels: 7

DB: 1 Gaps: 1

US-10-602-220-16 (1-417) x AL554978 (1-1086)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20

Db 6 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGC-CTTCGCTTCTCGGCC 64

Qy 21 LeuValSerTrrPhePleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40

Db 65 CTCGTTTCTCGGGACATCCCTGGGGCTAGAGCACTGGCAATGATGGCAAGGACGCT 124

Qy 41 ThrMetGlyTrrLeuHisTrrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60

Db 125 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 184

Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80

Db 185 GATTCTCTGCATCAGTGAGAAGCTTTCATGAGAGATGGCAGAGCTCATGGTCTCAGAAGGC 244

Qy 81 TrpIleAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrrMetAlaProGlnArg 100

Db 245 TGGAAGGATCAGGTTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 304

Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120

Db 305 GATTCAGAAGCGACACTTCAGGCAGACCCCTCAGCGCTTCTCATGGGATTCGCAGCTA 364

Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 365 GCTAAATTATGTTACACAAAGGACTGAAGTAGGATTTATGCAGATGTTGGAAATATAA 424

Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrrAspIleAspAlaGlnThrPheAla 160

Db 425 ACCTGGCCAGGCTTCCCTGGGAGTTTGGATACCTACGACATTAAGTGGTGGTGGTGGT 484

Qy 161 AspTrrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

Db 485 GACTGGGAGTAGATCTGCTTAAATTTGATGGTTGTTACTGTGACAGTTTGGAAATTTG 544

Qy 181 AlaAspGlyTrrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200

Db 545 GCAGATGGTTATAAGCACATGTCCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 604

Qy 201 SerCysGluTrrProLeuTrrMetTrrProPheGlnLysProAsnTrrThrGluIleArg 220

Db 605 TCCTGTGAGTGGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATATACAGAATTCGA 664

Qy 221 GlnTyrCysAsnHisTrrArgAsnPheAlaAspIleAspAspSerTrrLysSerIleLys 240

Db 665 CAGTACTGCATCACTGGCGGAAATTTTGTGCTGACATTCATGATTCCTGGAAGATATAAG 724

Qy 241 SerIleLeuAspTrrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

Db 725 AGTATCTTGGACTGGACATCTTTTAAACAGAGAGAAATTTGTGATGTTGTCTGGACAGGG 784

Qy 261 GlyTrrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrrAsnGlnGln 280

Db 785 GGTGGAAATGACCCAGATATGTTAGTATGGCAACTTTTGGCCCTCAGCTGGAATCAGCA 844

Qy 281 ValThrGlnMetAlaLeuTrrAlaIle-MetAlaAlaProLeuPheMetSerAsnAspLe 300

Db 845 GTAACTCAGATGGCCCTCTGGGSTATTTTCATGGCTGCTCCTTTATTCATGCTAATGACT 904

```

Qy 300 uArgHisIleSerProGlnAlaIysAlaLeuLeuGlnAspLysAspValIleAlaIleAs 320
Db 905 CCACACATCAGCCCTCAGCCCAAGCTCTCCCTTCAGGATPAGGACGTWATTGCCATCAA 964

Qy 320 nGlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlnGlnGlnGlnGln 340
Db 965 TCAGGACCCYTK-GGCAAGCAAGGTTACAGCTTAGAWAGGGRGACAM-TTTCWAGTGTG 1022

Qy 340 pGluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGln 357
Db 1023 GGA-MGA---CTCTCTCAGGTTAGCTGGGCTAGY-ATGATAAACCGGAGG 1069

RESULT 9
BX354096 1014 bp mRNA linear EST 23-APR-2004
LOCUS BX354096 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
DEFINITION cDNA clone CS0DC014YH04 5-PRIME, mRNA sequence.
ACCESSION BX354096
VERSION BX354096.2 GI:46550107
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1014)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 5, 2003 this sequence version replaced gi:30379845.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DC014DD02QP1&c=10506.r.
FEATURES
Location/Qualifiers
1..1014
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC014YH04"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN
Alignment Scores:
Pred. No.: 1.69e-175 Length: 1014
Score: 1699.00 Matches: 317
Percent Similarity: 97.54% Conservative: 0
Best Local Similarity: 97.54% Mismatches: 7
Query Match: 74.10% Indels: 4
DB: 5 Gaps: 0

US-10-602-220-16 (1-417) x BX354096 (1-1014)

Qy 5 AsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAlaLeuValSerTrp 24
Db 31 AACCAAGACTACATCTGGGCTGGCGCTTCGCSST-CCCTTCCTGGCCCTCGTTCTCTGG 89

Qy 25 AspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrp 44
Db 90 GACATCCCTGGGCTAGACACTGGCAATGGATTGGCAAGGACGCTACCATGGGCTGG 149

```

```

Qy 45 LeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluProAspSerCysIle 64
Db 150 CTGCATCTGGAGGCGCTTCATGTCACACCTTGATGCGCAGGAAGCCAGATTCCTGCATC 209

Qy 65 SerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAla 84
Db 210 AGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGCTCTCAGAAGGCTGGAAGGATGCA 269

Qy 85 GlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGly 104
Db 270 GGTATTAGTACCTCTGCAATTGATGATGCTGCTCCCAAGAGATTCAGAAGGC 329

Qy 105 ArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeuAlaAsnTrpVal 124
Db 330 AGACTTCAGGCAGACCCCTCAGCCCTTCTCATGGGATTCGCCAGCTAGCTAATTATGTT 389

Qy 125 HisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLysThrCysAlaGly 144
Db 390 CACAGCAAGGACTGGAAGCTAGGGAATTTATGCAGATGTTGGAAATAAAACCTGCGCAGGC 449

Qy 145 PheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAlaAspTrpGlyVal 164
Db 450 TTCCTCTGGGAGTTTGGATACACGACATTCATGCCAGACCTTGTCTGACTGGGGAGTA 509

Qy 165 AspLeuLysPheAspGlyCysTyrCysAspSerLeuGlnAsnLeuAlaAspGly-TY 184
Db 510 GATCTGCTAAAATTTGATGTTGTTACTGTGCACAGTTTGGAAAATTTGGCAGATGGTATA 569

Qy 184 rLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyrSerCysGluTr 204
Db 570 TAAGCACATGCTCTTGGCCCTGAATAGGACTGCGCAGAAAGCATTTGTGTACTCTCTGTGAGTG 629

Qy 204 pProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArgGlnTyrCysAs 224
Db 630 GCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTATACAGAAATTCGACAGTACTGCAA 689

Qy 224 nHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAs 244
Db 690 TCACTGGCGAAATTTTGTGACATTTGATGATTCTCGGAAAAGTATAAAGAGTATCTTGA 749

Qy 244 pTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAs 264
Db 750 CTGCACATCTTTTAAACCAGGAGAGAAATTTGTATGTTGCTGTCGACCCAGGGGTTGGAATGA 809

Qy 264 pProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGlnValThrGlnMe 284
Db 810 CCAGATATGTTAGTATGGCAACTTTGGCCCTCAGCTGGAATCAGCAAGTAATCTCAGAT 869

Qy 284 tAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSe 304
Db 870 GGGCTCTGGGCTATCATGCTGCTCTTTTATTCATGCTTAATGACCTCCGACACATCAG 929

Qy 304 rProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsnGlnAspProLe 324
Db 930 CCCCAAGCCAAAGCTCTCCTTCAGGRTAAGRCGTAAATTTGCCAC-AAWCAGGACCCCTK 988

Qy 324 uGlyLysGlnGly 328
Db 989 -GGCMAGCAAGGK 1000

RESULT 10
AL552630 999 bp mRNA linear EST 30-MAR-2004
LOCUS AL552630 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION clone CSODI067YJ01 5-PRIME, mRNA sequence.
ACCESSION AL552630
VERSION AL552630.3 GI:45857411
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Hominidae; Homo.
1 (bases 1 to 999)
Li, W.B., Gruber, C., Jesses, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 15, 2001 this sequence version replaced gi:31274445.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?r=CS0DI067CE01QPl&c=10506.r.

FEATURES
source

1. 999
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI067YJ01"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

Alignment Scores:

Pred No.:	1.11e-174	Length:	999
Score:	1691.50	Matches:	323
Percent Similarity:	98.48%	Conservative:	0
Best Local Similarity:	98.48%	Mismatches:	5
Query Match:	73.77%	Indels:	5
DB:	1	Gaps:	0

US-10-602-220-16 (1-417) x AL552630 (1-999)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 7 ATGCAGCTGAGGACCCAGCAATACATCTGGCTGCGGCTTGGCTT-CGCTTCTGGCC 65
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspGlnGlyLeuAlaArgThrPro 40
Db 66 CTGCTTTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTGGCAAGACCGCT 125
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 126 ACCATGGGCTGGCTGGCTGGGACGGCTTATGTGCACTTGTCTCCAGGAGAGCCA 185
Qy 61 AspSerCysIleSerGluLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 186 GATTCCTGCATCAGTGAAGCTCTTCTATGAGATGCGAGCTCATGGTCTCAGAAGGC 245
Qy 81 TrpIleAspAlaGlyTrpGluTrpLeuCysAlaLeuAspCysTrpMetAlaProGlnArg 100
Db 246 TGGGAAGATGAGGTTATGATGATCTCTGCACTTGTGATGCTTGGATGGCTCCCAAGA 305
Qy 101 AspSerGluArgGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 306 GATTCAGAGGACACTTCAGGCGAGACCTCAGGCGCTTCTCTATGGGATTCGCAGCTA 365
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 366 GCTAATATGTTACAGCAAGAGGACTGAGCTAGGATGATGATGATGATGATGATGATGAT 425
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 426 ACCTGCGCAGGCTTCCCTGGGAGTTTTGGATCTACGACATTTGATGCTCCAGACCTTTGCT 485

Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 486 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGCTTTGGAAATTTG 545
Qy 181 AlaAspGlyTrpIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 546 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAGCAATTTGTGAC 605
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 606 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 665
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
Db 666 CAGTACTGCAATCACTGGCGAAATTTTCTGCTGACATTCATCTCTGCGAAAGTATTAAG 725
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 726 AGTATCTTGGACTGGACATCTTTAAACCAGAGAGAAATTTGTATGTCTGGACCAAGG 785
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 786 GGTGGATGATCCAGCATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGAATCAGCAA 845
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 846 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTCTCT-TTATTCATGTCTAATGACCTC 904
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 905 CGACATCATCAGCCCTCAAGCCAAGCT-CTCCTTTCAGGATAAGGACGTAATTTGCAWC-AAT 962
Qy 321 GlnAspProLeuGlyLysGlnGly 328
Db 963 CAGACCCCTTG-GCAGCAAGAGGK 985

RESULT 11
BM564282

LOCUS
DEFINITION
5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE

1 (bases 1 to 1067)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.

Email: cga@nci.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM12759 row: p column: 07

High quality sequence stop: 761.

Location/Qualifiers

1. 1067

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5742198"

/tissue_type="medulla"

/lab_host="DH10B"

FEATURES

source

Best Local Similarity: 95.71% Mismatches: 5
Query Match: 72.37% Indels: 6
DB: 3 Gaps: 1

US-10-602-220-16 (1-417) x BQ062192 (1-1005)

Qy 28 GlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrProThrMetGlyTrpLeuHisTrp 47
Db 1 GGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCTACCAATGGCTGGCTGGCTGCACTGG 60

Qy 48 GluArgPheMetCysAsnLeuAspCysGlnGluProAspSerCysIleSerGluLys 67
Db 61 GAGCGCTTCATGTGCAACCTTGACTCCAGAGAGCCAGATTCCTCATCAGTGAGAAG 120

Qy 68 LeuPheMetGluMetAlaGluLeuMetValSerGluGlyTrpLysAspAlaGlyTrpGlu 87
Db 121 CTCTTCATGAGATGGCAGAGCTCATGCTCAGAAGCTGGAAGGATGCGAGGTATGAG 180

Qy 88 TyrLeuCysIleAspAspCysTrpMetAlaProGlnArgAspSerGluGlyArgLeuGln 107
Db 181 TACCTCTGCATTGATGACTGTGGATGGCTCCCAAGAGATTTCAGAAGGCGAGACTTTCAG 240

Qy 108 AlaAspProGlnArgPheProHisGlyLeuArgGlnLeuAlaAsnTrpValHisSerLys 127
Db 241 GCAGACCTCAGCGCTTTCCTCATGGATTCGCAGCTAGCTAATTAATGTTTCAGCAAA 300

Qy 128 GlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLysThrCysAlaGlyPheProGly 147
Db 301 GGACTGAAGCTAGGATTTATGAGATGTTGGAATTAACCTGGCGAGGCTTCCCTGGG 360

Qy 148 SerPheGlyTrpTrpAspIleAspAlaGlnThrPheAlaAspTrpGlyValAspLeuLeu 167
Db 361 AGTTTGGATACACGACATTGATGCCAGACCTTTCCTGACTGGGAGTAGATCTGCTA 420

Qy 168 LysPheAspGlyCysTrpCysAspSerLeuGlnLeuAlaAspGlyTrpLysHisMet 187
Db 421 AAATTTGATGTTGTTACTGTGACAGTTGGAATTTGGCAGATGGTTAAGACATG 480

Qy 188 SerLeuAlaLeuAsnArgThrGlyArgSerIleValTrpSerCysGluTrpProLeuTyr 207
Db 481 TCCTTGGCCCTGATAGACTGGCAGAGCAATGTGTACTCTCTGTGATGGCTCTTTAT 540

Qy 208 MetTrpProPheGlnLysProAsnTrpThrGluIleArgGlnTrpCysAsnHisTrpArg 227
Db 541 ATGTGGCCCTTCAAAGCCCAATTATACAGAAATCCGACAGTACTGCAATCACTGGCGA 600

Qy 228 AsnPheAlaAspIleAspAspSerTrpLysSerIleLysSerIleLeuAspTrpThrSer 247
Db 601 AATTTGCTGACATTGATGATTCCTGGAAAGTATAAAGAGATATCTTGGACTGGACATCT 660

Qy 248 PheAsnGlnGluArgIleValAspValAlaGlyProGlyGlyTrpAsnAspProAspMet 267
Db 661 TTTAACCAAGGAGAGATTTGATGTTGCTGGACAGGGGTTGGATGATGATCCAGATATG 720

Qy 268 LeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGlnValThrGlnMetAlaLeuTrp 287
Db 721 TTAGTGATGGCACTTGGCCCTCAGCTGGAATCAGCAAGTAATCACTCAGATGGCCCTCTGG 780

Qy 288 AlaIleMetAlaAlaProLeuPheMetSerAsnAspLeuArgHisIleSerProGlnAla 307
Db 781 GCTATCATGCTGCTCTCTTTATCATGTCTAATGACCTCCGACACATCAGCCCTCAAGCC 840

Qy 308 LysAlaLeuLeuGlnAspLysAspValIleAla-IleAsnGlnAspProLeu-GlyLysG 327
Db 841 AAAGCTCTTCTCAGGATTAAGAGCTTATTTGCCCATCAATCAGGACCCCTTGGGGCAAGC 900

Qy 327 In-GlyTrpGlnLeu-ArgGlnGly-AspAsnGlu---ValTrpGluArgProLeuS 345
Db 901 AAAGTTTACCAGCTTTAAACAGGGGAAACCACTTTGAAGGGTGGGAAACGACCCCTCTCT 960

Qy 345 erGlyLeu 347
Db 961 CAGGGCTA 968

RESULT 13
CO645623
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

CO645623 922 bp mRNA linear EST 23-JUL-2004
ILLUMIGEN MCQ 24642 Katze MMLV Macaca mulatta cDNA clone
TBIUM:23923 57 similar to Bases 5 to 920 highly similar to human
GLA (H8.69089), mRNA sequence.
CO645623 GI:50567117
EST.
Macaca mulatta (rhesus monkey)
Macaca mulatta
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Cercopithecoidea; Cercopithecinae; Macaca.
1 (bases 1 to 922)
Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., AGY, M.B.,
Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
Tadonato, S.P.
Analysis of the Macaca mulatta transcriptome and the sequence
divergence between Macaca and human
Genome Biol. 6 (7), R60 (2005)
15998449
Contact: C. Magness
Illumigen Biosciences Inc.
2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
Tel: 2063780400
Fax: 2063780408
Email: cmagness@illumigen.com
Sequenced on 2004.03.02. 796 Q20 bases. Library Preparation: Prof.
Michael Katze lab at University of Washington DNA Sequencing:
Illumigen Biosciences Inc. For further information, see
http://www.macaque.org
PCR Primers
FORWARD: CCCTCACTAAAGGGACAAAA
BACKWARD: CACTATAGGCGAATTGGTA
Insert length: 922 Std Error: 0.00
Plate: CL000111 row: E column: 08
Seq primer: CCCTCACTAAAGGGACAAAA
POLYA=No.

FEATURES
source
1..922
/organism="Macaca mulatta"
/mol_type="mRNA"
/strain="Indian"
/db_xref="taxon:9544"
/clone="TBIUM:23923"
/sex="female"
/dev_stage="adult"
/lab_host="Electromax DH10B"
/clone_lib="Katze MMLV"
/notes="Organ: liver; Vector: pDONR 222; Site 1: BarG I;
Site 2: BarG I; Created from CloneMiner cDNA Library
Construction kit (catalog #18249-029)"

ORIGIN
Alignment Scores:
Pred. No.: 4,59e-170 Length: 922
Score: 1649.00 Matches: 295
Percent Similarity: 98.03% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 6
Query Match: 71.91% Indels: 0
DB: 7 Gaps: 0

US-10-602-220-16 (1-417) x CO645623 (1-922)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 11 ATGCAGCTGAGGAACCCAGAAATGCATCTGGCTGCGCTTCTCTTCTCTGCGC 70

Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 71 CTCGTTTCTCTGGGACATCCCTCTGGGGCCAGACACTGGCAATGATTTGGCAGACGCT 130

JOURNAL
COMMENT

Unpublished (2001)
On Feb 16, 2001 this sequence version replaced gi:31315840.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
10506.r

For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DK009BC06NFI&c=10506.r.

FEATURES

source

1..1026
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009VE12"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
Pred. No.: 6.14e-170 Length: 1026
Score: 1648.50 Matches: 308
Percent Similarity: 98.41% Conservative: 1
Best Local Similarity: 98.09% Mismatches: 5
Query Match: 71.89% Indels: 2
DB: 1 Gaps: 0

US-10-602-220-16 (1-417) x AL577581 (1-1026)

Qy	97	AlaProGlnArgAspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGly	116
Db	955	GCTCCCCAAAGAGATTGAGAACG-AGACTTCAGCA-GACCCCTCAGCCCTTTCCTCATGGG	898
Qy	117	IleArgGlnLeuAlaAenTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAsp	136
Db	897	ATTGCCAGCTAGCTAATTATGTTACAGCAAGGACTGAAGCTAGGATTTATGTCAGAT	838
Qy	137	ValGlyAsnLysThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAla	156
Db	837	GTTGGAAATAAACCTCGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTGATGCC	778
Qy	157	GlnThrPheAlaAspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSer	176
Db	777	CAGACCTTTGCTGACTGGGAGTAGATCTGCTAAATTTTGGTGGTTGTTGCTGTGACAGT	718
Qy	177	LeuGluAsnLeuAlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArg	196
Db	717	TTGGAAATTTGGCAGATGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGA	658
Qy	197	SerIleValTyrSerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyr	216
Db	657	AGCATTTGTACTCTCTGAGTGGCCCTCTTTATATATGTCGCTTTTCAAAAGCCCAATTAT	598
Qy	217	ThrGluIleArgGlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyr	236
Db	597	ACAGAAATCCGACAGTACTGCAATCTGCGGAAATTTTGTCTGACATTGATGATCTCTGG	538
Qy	237	LysSerIleLysSerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspVal	256
Db	537	AAAGTATAAGAGTATCTTGGACTGACATCTTTTAAACAGGAGAGAGAAATTTGTTGATGTT	478
Qy	257	AlaGlyProGlyGlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSer	276

Qy	41	ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	131	ACATGGGCTGCTGCACTGGAGCGCTTCATGTCACCTTGACTCCAGAGAGACCA	190
Qy	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	191	GATTCCTGCATCAGTGAGAGCTCTTCATGAGATGCGAGAGCTCATGGTCTCAGATGGC	250
Qy	81	TripLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg	100
Db	251	TGGAAGGATCAGGTTATGATGATCTCTGATGATGATGATGATGATGATGATGATGATGAT	310
Qy	101	AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu	120
Db	311	GATTTAGAGGACAGACTTCAGGAGAGCCCTCAGGCGCTTTCTCATGGGATTCGCCAGCTA	370
Qy	121	AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys	140
Db	371	GCTAATTTATGTTACACCAAGGAGCTGAGCTAGGATTTATGACAGATTTGGAATATAA	430
Qy	141	ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla	160
Db	431	ACCTGCCAGGCTTCCCTGGGAGTTTGGATCTATGACATTTGATGCCAGACCTTTGCT	490
Qy	161	AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu	180
Db	491	GACTGGGAGTAGATCTGCTGAAATTTGATGTTGTTTACTGTGACATTTGGAAAGTTG	550
Qy	181	AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr	200
Db	551	GCAGATGTTATAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGTAC	610
Qy	201	SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg	220
Db	611	TCCTGTGAGTGGCTCTTTATATGTCGCAATTTCAAAGGCCCAATTACACAGAAATCCGA	670
Qy	221	GlnTyrCysAsnHisTyrArgAsnPheAlaAspIleAspAspSerTyrLysSerIleLys	240
Db	671	CAGTACTGCAATCACTGGCGAAATTTTGTGACATGATGATCTCTGCGAAAGTTAAG	730
Qy	241	SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly	260
Db	731	AGTATCTTGACTGGACATCTTTAACAGGAGAGATTTGTGATGTTGCTGGACACAGG	790
Qy	261	GlyTyrAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTyrAsnGlnGln	280
Db	791	GGTTGGAATGACCCAGATATGTTAGTGAATTGGCAACTTTGGCCCTCAGCTGGAATCAGCA	850
Qy	281	ValThrGlnMetAlaLeuTyrAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
Db	851	GTAACCTCAGATGGGCCCTCTGGCTATCATGCTCTCTCTTTATTCATGCTTAATGACCTC	910
Qy	301	ArgHisIleSer	304
Db	911	CGACACATCACC	922

RESULT 14

AL577581/c

LOCUS

DEFINITION AL577581 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DK009VE12 3-PRIME, mRNA sequence.

ACCESSION

AL577581

VERSION

AL577581.3

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 1026)

Li, W.B., Gruber, C., Jesses, J. and Polayes, D.

Full-length cDNA libraries and normalization

```

Db 477 GCTGGACACGGGGTTGGAAATGACCCAGATATGTTAGTATGGCAACTTTGGCCTCAGC 418
Qy 277 TTPAsnGlnValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMet 296
Db 417 TGGAAATCAGCAAGTAACCTCAGATGGCCCTCTGGGGTATCATGGCTGCTCTTTATTTCATG 358
Qy 297 SerAsnAspLeuArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspVal 316
Db 357 TCTAATGACCTCCGACACATCAGCCCTCAAGCCAAAGCTCTCTTACAGGATAAGGACGTA 298
Qy 317 IleAlaIleAsnGlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsn 336
Db 297 ATTGCCATCAATCAGGACCCCTTGGCAAGCAAGGGTACCAGCTTAGACAGGAGACAC 238
Qy 337 PheGluValTrpGluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArg 356
Db 237 TTTGAAGTGTGGGAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGTATGATAAACCGG 178
Qy 357 GlnGluIleGlyProArgSerThrIleAlaValAlaSerLeuGlyLysGlyVal 376
Db 177 CAGGAGATTGGTGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAAGGAGTG 118
Qy 377 AlaCysAsnProAlaCysPheIleThrGlnLeuLeuProValLysArgGlyLysLeuGlyPhe 396
Db 117 GCCTGTAAATCTGCTGCTCTCATCACACAGCTCTCTCMTGTGAAAGGAAGCTAGGGTTC 58
Qy 397 TyrGluTrpThrSerArgLeuArgSerHisIleAsnProThr 410
Db 57 TATGATGGACTTCAGGTTAAGAGTAAACATAATVCRACA 16

RESULT 15
AY408542 1296 bp DNA linear GSS 15-DEC-2003
LOCUS Mus musculus GLA gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY408542
VERSION AY408542.1 GI:39764513
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1296)
Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
    source          Location/Qualifiers
    1..1296
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    <1..>1296
    /gene="GLA"
    /locus_tag="HMC3258"

gene

ORIGIN
Alignment Scores:

```

```

Pred. No.: 2,02e-167 Length: 1296
Score: 1627.00 Matches: 300
Percent Similarity: 81.06% Conservative: 38
Best Local Similarity: 71.94% Mismatches: 79
Query Match: 70.96% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x AY408542 (1-1296)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGAAGCTTTTGACGACAGATACGCGCTGGTCTGTGAGCTTTGGCTTTGTCCTCGCT 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 TTAGTTTCTCGAGCATCTCTGGGCTCAGAGCATTTGACCAATGGCTTTGGCGGACCTCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACTATGGGCTGGCTGCAATTGGGAACGTTTCATGTGCAACCTTGACTCCCAAGAGAGCCT 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATGCTGCATAGTAGGCAACTGTTTCATGCAGATGCGCAGGCTCATGGTCTCTGATGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGCGGGATGACGTTATGACTATCTCGCATAGTAGACTGTTGGATGGCTCCCGAGAGG 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAAGGGCAGGCTTCAGGCGAGATCCCAACGCTTCTTAGTGGGATCAACACCTC 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCAATTCGTCACAGCAAGGATTTGAAGTAGAGGATTTATGCAGATGTTGGGAATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGTGACGTTTCCCGGAGTTTGGATCTCTATGACATTCATGCGCAGACATTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGCGTAGATCTGCTAAATTTGATGGTTGTGCTGACAGTGTAGTATCTCTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GAGAAATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 TATTACTGCAATCATTTGGAGAAATTTTATGATGATTTTATGATCTCTCGGGAAGCATAAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AATATCTTGTCTTGGACAGTGGTTTACCAGAGAGATTTGTTGAGTCTGCTGGACACGAGC 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 AGCTGGAAATGACACACATGTTAGTGTATGATCGGCAACTTTGGCCTCAGTTGGACACGAG 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTGACACAGATGGCCCTCTGGGCTATCATGCGCGCTCTCTCTCATGTCATGTCACAGATCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGCAAAATCAGCTCTCAGCCAAAGCTCTCTCTCAGAAATAGGATGTAATTCATTCATC 960

```

```
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
961 CNAGACCCCTTGGCGAAGCAGGGCTACTGTTTCAGAAAGGAAACACCATTTGAGGTTTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1021 GNACGGCCACTCTCCAACTAGCCTGGGCTGGGCTGTGAGAAACCTGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1081 GGACCTTTGTCCTTATACCATCCAGATTCTTCCCTGGGTAGAGGACTAGCCTGCAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1141 GGCTGCATCATTAATCAGCTTCTCCCGAGNAAGTACACCTAGGCTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1201 TTGACCTTAAAAAACTCGAGTAACCCCTCAGGCACCTGTTTGTTCGGTTA 1251
```

Search completed: January 1, 2006, 05:47:02
Job time : 7590.48 secs

PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KUMAI/) KUMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 XX
 DR WPI: 2002-681656/73.
 DR P-PSDB: A828212.
 XX

PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 PT storage diseases.
 XX

PS Claim 1; Page 47-49; 88pp; English.
 XX

CC The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-12 DNA
 XX

SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,34e-235 Length: 1254
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x A4D45225 (1-1254)

QY 1 MetGlnLeuArgAnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGGTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLeuHisLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCCTGCATCAGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAAAGATGCAAGTTATGATGATACCTTCGATTTGATGACTTTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTCAGAGCGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAAATATGTTACAGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCGAGCTTCCTCGGAGATTTGGATATCTACGACATTTGATGCCAGACTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGATCTGCTAAAAATTTGATGGTGTGTACTGTGACAGTTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGCACATGTCCTTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATATGTCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCAGTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAGATATAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTAICTTGGATGGACATCTTTTAACAGAGAGAGAAATTTGTGATGTTGCTGGACACGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTGTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGGCAAGCAAGGGTACAGCTTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGAGCTATGATAAACCGCAGGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAAGAGTGGCTGTATCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTGTCTTCAGCTA 1251
 RESULT 2
 ADD84754
 ID ADD84754 standard; DNA; 1254 BP.
 XX
 AC ADD84754;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-12 DNA.
 XX
 KW Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1254
 FT /*tag= a
 FT /product= "Human rGAL-12"
 XX
 PN US2003106095-A1.
 XX
 PD 05-JUN-2003.
 XX

PF 20-MAR-2002; 2002US-00103327.
XX
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
PI
XX
XX WPI; 2003-801257/75.
DR P-PSDB; ADD84755.
XX
XX
PT New polynucleotide for producing active recombinant human and animal
PT lysosomal enzymes in a plant expression system that can be used in enzyme
PT replacement therapy.
XX
XX
XX Claim 1; SEQ ID NO 15; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
CC nucleic acids encoding them. The polypeptides are used in a method for
CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents DNA encoding a human alpha-
CC galactosidase derivative polypeptide of the invention.
XX
XX Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;
SQ

Alignment Scores:
Pred. No.: 2,34e-235 Length: 1254
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x ADD84754 (1-1254)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTCGCGCTTCGCTTCCTGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGCTTCTCGGACATCCCTGGGCTAGAGCACTGGACATGGATGGCAAGCAGCCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGGCCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCATGTAGAGAGCTTCATGTGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGAGATGCAGGTATGATGACCTCTGATGTAGTACTGTTGGATGGCTCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyLeuArgGlnLeu 120
DB 301 GATTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTCCTCATGGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTTCACAGCAAGAGACTCAAGCTAGGATTTATGCAGATGTGGAAATAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGCGCTTCCTGGAGATTTTGGATATACAGACATGTATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180

DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGCACAGTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGACGATTTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluLeuArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTCAAAAGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACCTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAGATATAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTGTGGACCAAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGAAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATGGCATCAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGGCAAGCAAGGTTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCCTGGCTGTAGCTATGATAAACCCGCGAGAGTTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGCATTTGCTTCCTGGGTAAAGAGTGGCTGTATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCTCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTCACATNAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 3
ADM48684
ID ADM48684 standard; DNA; 1254 BP.
XX
AC ADM48684;
XX
DT 03-JUN-2004 (first entry)
XX
XX Human wild type rGAL-12 DNA.
XX
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FT 1. .1254
FT /*tag= a
FT /product= "GAL-12 protein"
XX

PN US2004023281-A1.
 XX
 PD 05-FEB-2004.
 XX
 PF 23-JUN-2003; 2003US-00602220.
 XX
 PR 26-FEB-1988; 88US-00160766.
 PR 26-FEB-1988; 88US-00160771.
 PR 17-FEB-1989; 89US-00310881.
 PR 22-OCT-1990; 90US-00600244.
 PR 31-JUL-1992; 92US-00923692.
 PR 30-DEC-1992; 92US-00997733.
 PR 28-DEC-1993; 93US-00176414.
 PR 19-JAN-1994; 94US-00184237.
 PR 14-OCT-1994; 94US-00324003.
 PR 21-MAY-1999; 99US-00316572.
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX
 (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 PA (POGU/) POGUE G P.
 PA (ERWI/) ERWIN R L.
 PA (GRIL/) GRILL L K.
 XX
 Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
 PI
 DR WPI; 2004-142650/14.
 DR P-PSDB; ADM48685.
 XX
 New alpha-galactosidase polypeptides, useful in producing recombinant
 PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
 PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
 PT disease.
 XX
 Disclosure; SEQ ID NO 15; 72pp; English.
 PS
 XX The present invention relates to novel galactosidase (Gal) proteins such
 CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
 CC of the present invention are useful for producing recombinant lysosomal
 CC enzymes for enzyme replacement therapy for treating human and animal
 CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
 CC human wild type rGAL-12 DNA used in the exemplification of the invention.
 XX
 SQ Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,34e-235 Length: 1254
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0
 US-10-602-220-16 (1-417) x ADM48684 (1-1254)
 QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGCTGGCGCTTGGCTTCCTCGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGCTTTCCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
 Db 121 ACCATGGGCTGGCTGCACTGGAGCGCTTCATGTGCACACCTTGACTCCAGGAGAGACCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCTGCATCAGTAGAAGACTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAAAGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGNAGGATGCAGGTTATGAGTACCTCTGCAATTGATGACTGTGTGGATGCTCCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCCTCATGGGATGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGAGCTTCCCTGGGAGTTTGGATACACTACGACATTGATGCCCAGACCTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGACTCTGCTAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAATTTG 540
 QY 181 AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyVArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATGTGTAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCTGTGAGTGGCTCTTTATATGTCGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTCTGCATTTGATGATTCCTGGGAAAGTATAAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACCAGAGAGAAATTTGTGATGTTGCTGCACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 Db 781 GGTGTGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
 QY 301 ArgHisLeuSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGACGTAATTGCCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGlyValTrp 340
 Db 961 CAGGACCCCTTGGCAAGCAAGGTTACCACTTAGACAGGAGACAACCTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGGCTTGGGCTGTAGCTATGATNAACCGCGCAGGAGATTGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAserLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGAACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGAGTGGCTGTAACTCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCCTATGAATGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGTTAAGAAAGTCACATAAAATCCCAAGGCACCTGTTTGTCTTCAGCTA 1251
 RESULT 4
 ADU66919 standard; DNA; 1254 BP.
 XX
 AC ADU66919;


```

Db      1201 TCAGGTTAAGAAGTCACATAAATCCACAGGCACGTGTTTGGTTTCAGCTA 1251
RESULT 5
ID      AEA27448
XX      AEA27448 standard; DNA; 1254 BP.
XX      AEA27448;
XX
XX      11-AUG-2005 (first entry)
XX      Human alpha-galactosidase DNA, rGAL-12, SEQ ID NO: 15.
XX
XX      Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX      genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX      antileptic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX      gene; ds.
XX
XX      Homo sapiens.
XX
XX      Key      Location/Qualifiers
XX      CDS      1..1254
XX              /*tag= a
XX              /product= "Human alpha-galactosidase protein"
XX
XX      US2005125859-A1.
XX
XX      09-JUN-2005.
XX
XX      08-NOV-2004; 2004US-00984389.
XX
XX      26-JUL-2000; 2000US-00626127.
XX      13-NOV-2001; 2001US-00993059.
XX      20-MAR-2002; 2002US-00103327.
XX
XX      (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX      Garger SJ, Turpen TH, Kumagai MH;
XX      WPI; 2005-404004/41.
XX      P-PSDB; AEA27449.
XX
XX      New isolated polypeptides useful for producing lysosomal enzymes in
XX      plants to be utilized in enzyme replacement therapy or for the
XX      therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX      Gaucher's disease.
XX
XX      Disclosure; SEQ ID NO 15; 88pp; English.
XX
XX      The present invention relates to the production of human and animal
XX      lysosomal enzymes in plants by a transient plant expression system. The
XX      invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX      galactosidase (Gal) enzymes having a post-translational modification
XX      provided by the plant expression system. The invention is useful in
XX      enzyme replacement therapy for treating lysosomal storage diseases such
XX      as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX      disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX      also useful in researches for developing new approaches to medical
XX      treatment of lysosomal storage diseases and in industrial processes
XX      involving enzymatic substrate hydrolysis. The present sequence is the
XX      human alpha-galactosidase DNA.
XX
XX      Sequence 1254 BP; 316 A; 289 C; 328 G; 321 T; 0 U; 0 Other;
XX
XX      Alignment Scores:
XX      pred. No.:      2,34e-235      Length:      1254
XX      Score:          2293.00      Matches:      417
XX      Percent Similarity: 100.00%      Conservative: 0
XX      Best Local Similarity: 100.00%      Mismatches: 0
XX      Query Match:      100.00%      Indels:      0
XX      DB:              14      Gaps:      0
XX
XX      US-10-602-220-16 (1-417) x AEA27448 (1-1254)

```

```

1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
1 ATGCAAGCTGAGGAAACCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCTTGGCC 60
21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
61 CTCGTTTCTCTGGGACATCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGGACGCT 120
41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCACCTTGACTGCCAGGAAGAGCCA 180
61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
181 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
241 TGGNAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
301 GATTTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTTCTCTCATGGGATTCGCCAGCTA 360
121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
361 GCTAATTATGTTCAACAGCAAGAGCTGAAGCTAGGGATTTATGCAGATGTGGAAATAAA 420
141 ThrCysAlaGlyPheProGlySerPheGlyTyrTrpAspIleAspAlaGlnThrPheAla 160
421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACTACGACATTCATGCCAGACCTTTGCT 480
161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
481 GACTGGGGAGTAGACTGCTAAATTTGATGGTTGTACTGTACAGTTTGGAAATTTG 540
181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
541 GCNAGTGGTTATTAAGCACATGCTTGGCCCTGNAATAGGACTGGCAGAGCATTTGTGTAC 600
201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATTTGATGATTCCTGGAAAGATATAAG 720
241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
721 AGTATCTTGGACTGGACATCTTTTAACAGGAGAGAAATTTGTGATGTGTGGACCAAGG 780
261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
781 GGTGTGAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCCTCAGCTGGAATCAGCAA 840
281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTTCATGTCTAATGACCTC 900
301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATTTGCCATCAAT 960
321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
961 CAGGACCCCTTGGGCAAGCAAGGTCACAGCTTAGACAGGAGACAACTTTGAGTGTGG 1020
341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
1021 GAACGACCTCTCTCAGGCTTAGCGCTGTAGCTATGATTAACCCGCGAGGATTTGGT 1080

```

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValIleAlaCysAsnPro 380
 DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCTCGGTAAAGAGTGGCCCTGTAATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
 DB 1141 GCTTGCTTCATCACAGCTCCCTCTGTGTAAGAGAGCTAGGGTCTATAGTGGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 DB 1201 TCAAGGTGAAGAAGTCACATAAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 6

AAD45223
 ID AAD45223 standard; DNA; 1266 BP.

AC AAD45223;

DT 27-DEC-2002 (first entry)

DE Human rGAL-8 DNA.

XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
 KW therapeutic; rGAL-8; gene; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1266

FT /tag= a

FT /product= "Human rGAL-8 protein"

XX US2002088024-A1.

PN 04-JUL-2002.

XX 13-NOV-2001; 2001US-00993059.

XX 26-JUL-2000; 2000US-00626127.

XX (GARG/) GARGER S J.

PA (TURP/) TURPEN T H.

PA (KUMAI/) KUMAGAI M H.

XX Garger SJ, Turpen TH, Kumagai MH;

XX WPI: 2002-681656/73.

DR P-PSDB; AAE28210.

XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
 storage diseases.

XX Claim 1; Page 42-44; 88pp; English.

XX The invention relates to human alpha-galactosidase truncated at the
 CC carboxy terminus and the production of enzymatically active recombinant
 CC human and animal lysosomal enzymes. The invention is useful for producing
 CC lysosomal enzymes for treating lysosomal storage diseases, producing
 CC altered or mutated proteins, enzymatically active or otherwise, to serve
 CC as precursors or substrates for further in vivo or in vitro processing to
 CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-8 DNA

XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2 37e-235 Length: 1266

Score: 2293.00 Matches: 417

Percent Similarity: 100.00% Conservatives: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-602-220-16 (1-417) x AAD45223 (1-1266)
 QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTTCTGGGACATCTCTGGGCTAGAGCACTGGACATGATGATGGCAGGACGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGCATCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
 QY 61 AspSerCysIleSerGlnLysLeuPheMetGlnMetAlaGlnLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TGAAGGATGCAGGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAGAAGGCAGACTTCAGGCAGACCTTCAGCGCTTCCTCATGGATTCGCCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTTATGTTTCACAGCAAGGAGCTGAAGCTAGGAGATTTATGCAGATGTGGAAAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACTGGCGAGGCTTCCTGGGAGTTTGGATACTACGACATGTATGATGCCAGACTTTGCT 480
 QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTACTGTGACAGATTTGGAAAAAT 540
 QY 181 AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATAGCACATGCTTCCTGGCCCTGCAATAGACTGGCAGAGACTGTGTATC 600
 QY 201 SerCysGluTyrProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 DB 661 CAGTACTGCAATCAGCTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAAGTATAA 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTGGAGCTGGACATCTTTTAACAGAGAGAGAAATTTGTGATGTTCGTGACAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GATTGGATGACCCAGATATGTTAGTATGGCACTTTGGCTTCAGCTCGAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTCTTTTATCATGTCTAATGAC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAAATTCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
 DB 961 CAGGACCCCTTGGGCAAGCAAGGATACCAGCTTTAGACGGGAGACAACTTTGAAGTGTGG 1020

Db 961 CAGGACCCCTTGGCGAAGGCTACAGCTTAGACAGGAGACAACCTTTGAAGTGTG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAACCGCAGGAGATTGGT 1080
 Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCATGTTGCTTCCCTGGGTAAAGAGTGGCTGTAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAAGGAGCTTAGGGTTCTATGAATGGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAAGAGTCACATAAATCCACAGGCACCTGTTTGTCTTCAGCTA 1251
 RESULT 8
 ADJ88276
 ID ADJ88276 standard; DNA; 1266 BP.
 XX AC ADJ88276;
 XX DT 06-MAY-2004 (first entry)
 XX DE Human WT rGAL-8 (galactosidase) DNA.
 XX KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
 XX KW Fabry's disease; Gaucher's disease; human; gene; ds.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 XX FT CDS 1..1266
 XX FT /*cag= a
 XX FT /product= "Human WT rGAL-8 protein"
 XX PN US2004016021-A1.
 XX PD 22-JAN-2004.
 XX PF 23-JUN-2003; 2003US-00602219.
 XX PR 26-FEB-1988; 88US-00160766.
 XX PR 26-FEB-1988; 88US-00160771.
 XX PR 15-JUL-1988; 88US-00219279.
 XX PR 17-FEB-1989; 89US-00310881.
 XX PR 05-MAY-1989; 89US-00347637.
 XX PR 08-JUN-1989; 89US-00363138.
 XX PR 22-OCT-1990; 90US-00600244.
 XX PR 16-JAN-1991; 91US-00641617.
 XX PR 26-JUL-1991; 91US-00737899.
 XX PR 01-AUG-1991; 91US-00739143.
 XX PR 31-JUL-1992; 92US-00923692.
 XX PR 30-DEC-1992; 92US-00997733.
 XX PR 29-DEC-1993; 93US-00176414.
 XX PR 19-JAN-1994; 94US-00184237.
 XX PR 14-OCT-1994; 94US-00324003.
 XX PR 21-MAY-1999; 99US-00316572.
 XX PR 26-JUL-2000; 2000US-00626127.
 XX PR 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 XX (FOGU/) FOGUE G P.
 XX (ERWI/) ERWIN R L.
 XX (GRIL/) GRILL L K.
 XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
 XX WPI; 2004-108227/11.
 XX P-PSDB; ADJ88277.

PT New lysosomal enzymes, useful in treating human and animal lysosomal
 PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
 XX Claim 1; SEQ ID NO 11; 71pp; English.
 XX CC The invention relates to nucleotide encoding galactosidase (GAL). The
 CC invention is useful in gene therapy. The polynucleotides and polypeptides
 CC are useful in treating human and animal lysosomal storage diseases, e.g.
 CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
 CC DNA.
 XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 2,37e-235 Length: 1266
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 12 Gaps: 0
 US-10-602-220-16 (1-417) x ADJ88276 (1-1266)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGCTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACATGATGGATGGCAGACGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCA 180
 Qy 61 AspSerCysIleSerGlnIysLeuPheMetGluMetAlaGlnLeuMetValSerGluGly 80
 Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGGATGCAGGTTATGAGTACCTCTGGATCTGATGATGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTTCAGAGGCGAGACTTCAGGCGAGACCTTCAGCGCTTCTCATGGATTCGCCAGCTA 360
 Qy 121 AlaAsnTyrValHisSerIysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnIys 140
 Db 361 GCTAATTTATGTTCCACAGCAAGAGACTGAAGCTAGGAGATTTATGCAGATGTTGGAAATAAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCTGGAGTTTGGATATACGACATTCATGATGCCAGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLeuIysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGGAGTAGACTCTGCTAAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540
 Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCTGTGAGTGGCTCTTTATATGTGGCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGACATTCATGATGATCTCTGGAAAAGATATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260

```
Db 721 AGTATCTTGAGCTGGACATCTTTAAACACGAGAGAAATGTTGATGTTGCTGACCAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGAATGACCCAGATATGTTAGTGGTGGCAACTTTGGCCCTCAGCTGGAATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTNACTCAGATGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTTGGCAAGCAAGGGTATCCAGCTTAGACAGGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAGGAGTGGCCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGGCTTCAGCTA 1251
RESULT 9
ID ADM48680
XX ADM48680 standard; DNA; 1266 BP.
AC ADM48680;
XX
XX 03-JUN-2004 (first entry)
XX Human wild type rGAL-8 DNA.
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
XX lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
XX Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
XX gene; ds.
XX
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT FT /*tag= a
FT FT /product= "GAL-8 protein"
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
```

```
PR 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (FOGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48681.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
XX lysosomal enzymes for the treatment of lysosomal storage diseases, such
XX as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
XX disease.
XX
XX Disclosure; SEQ ID NO 11; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human wild type rGAL-8 DNA used in the exemplification of the invention.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-10-602-220-16 (1-417) x ADM48680 (1-1266)
Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGCCTTCCCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGTTTCTCTGGACATCTCTGGGGCTAGAGCACTGGACATGATGTGGCAAGCAGCCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGACCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTGTCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysAlaAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAGGATGCAGGTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTTCCACAGCAAGGACTGAAAGCTAGGATTTATGCATGTTGGAAATATAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTyxAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTTTGGATATACGACATTCATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
```



```
Db 481 GACTGGGGAGTAGATCTGCTAAAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTTG 540
Qy 181 AlaAspGlyTyrIysHisMetSerLeuAlaLeuAenArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTATAGACACATGCTCTGGCCCTGAATAGGACTGGCAGAACATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpIysSerIleLys 240
Db 661 CAGTACTGCAATCACTGCGGAAATTTTGTGACATTCATCTCTGGAAGATATAAG 720
Qy 241 SerIleLeuAspTrpTrpSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGTGCATCTTTTAAACGAGGAGAGATTTGTGATGTGTGGACACGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGACATATGTTAGTCAATGGCAACTTTGGCCCTCAGCTGGAATCAGAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTCAATGATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCCAAGCTCTCTTCAGGATTAAGGACGTAAATGGCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTrp 340
Db 961 CAGACCCCTTGGCAAGCAGGTTACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GlnArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGlnIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGGCTGTAGTATGATTAACCGCGAGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCATGTTGCTTCCCTGGGTAAGAGTGGCTGTATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACTGTTTTGTCTCAGCTA 1251

RESULT 10
ADU66915
ID ADU66915 standard; DNA; 1266 BP.
AC
XX
AC ADU66915;
XX
XX
DT 10-FEB-2005 (first entry)
XX
DE Human alpha-galactosidase protein encoding DNA #5.
XX
XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human; gene; ds.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..1266
XX /*tag= a
XX /product= "Human alpha-galactosidase protein"
```

```
PN US2004234516-A1.
XX
PD 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
PI
XX
XX WPI; 2004-821274/81.
DR
XX
XX P-PSDB; ADU66916.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 11; 86pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase enzymes having a post-translational modification provided
XX by the plant expression system. The invention is useful in enzyme
XX replacement therapy for treating lysosomal storage diseases such as
XX Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein encoding DNA.
XX
XX SQ Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-602-220-16 (1-417) x ADU66915 (1-1266)
Qy 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCTGGACATCCCTGGGCTAGAGACTGGACATGATGGATGGACAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTCAACCTTCACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGlnLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGTCATCATGTAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATGATGATACCTCTGCATTTGATGACTGTTGGTGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
```



```
Db 361 GCTAATATTGTCACAGCAAGGACTGAGAGCTAGGATTTATGACAGATGTTGGAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCTCGGAGTTTGGATACATGACGATGATGCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGATTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrIlyHisMetSerLeuAlaLeuAsnArgThrGlyVArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACACATGCTCTTGCCCTGATAGAGCTGGCAGAGCATTTGTGTC 600
Qy 201 SerCysGluTrpProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCAGCTGGCGAAATTTTGCTGCACATTTGATGATTCCTGGAAAGTATAAG 720
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGAGCTGGACATCTTTTAACAGGAGAGATTTGTTGATGTTGCTGGACGAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGAAATGACCCAGATATGTTAGTATTGGCAACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATCAGCCCTCAGCCAAAGCTCTCTCTCAGGATNAGGACGTAATGTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnGluValTrp 340
Db 961 CAGGACCCCTTGGCAGCAAGGCTACCACTTAGACAGGAGACAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCCTGGCTGTAGCTATGATGATAAACCGGCAGGAGATGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTCTTCTGGTAAAGAGTGGCTGTATCTCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACACAGCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCAATATAATCCACAGGCACTGTTTGTCTTCAGCTA 1251
```

RESULT 11

AEA27444

ID AEA27444 standard; DNA; 1266 BP.

AC AEA27444;

XX

XX

11-AUG-2005 (first entry)

DE Human alpha-galactosidase DNA, rGAL-8, SEQ ID NO: 11.

XX Gauchers disease; metabolic; neurologic disease; niemann pick disease;
KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
KW gene; ds.

XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT CDS 1..1266
FT /*tag= a
FT /product= "Human alpha-galactosidase protein"
XX
XX US2005125859-A1.
XX
XX 09-JUN-2005.
XX 08-NOV-2004; 2004US-00984389.
XX 26-JUL-2000; 2000US-00636127.
XX 13-NOV-2001; 2001US-00993059.
XX 20-MAR-2002; 2002US-00103327.
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
PI
XX
XX WPI; 2005-404004/41.
XX P-PSDB; AEA27445.
XX
XX New isolated polypeptides useful for producing lysosomal enzymes in
XX plants to be utilized in enzyme replacement therapy or for the
XX therapeutic treatment of human or animal lysosomal storage diseases, e.g.
XX Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 11; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase DNA.
XX
XX Sequence 1266 BP; 323 A; 291 C; 330 G; 322 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,37e-235 Length: 1266
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-10-602-220-16 (1-417) x AEA27444 (1-1266)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGTGCGGCTTCGCTTCCTCTCGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGACAATGGATGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGCTGGCTGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGCCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTTCATCAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240

```
QY 81 TtrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGGTTATAGTACCTCTGTCATTGATGACTGTGTGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGAGCAGACTTTCAGGCAGACCTTCAGGCGCTTCTCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLeuLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAGGACTGAAGCTAGGGATTTATGTCAGATGTTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCCAGCGCTTCCCTGGGAGTTTGGATACTACGACATTTGATGCCCGACGCTTTGCT 480
QY 161 AspTyrGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACATTTGGAAAATTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGTTTATAAGCACATGCTCTGGCCCTGGAATAGGACTGCGCAGAACATTTGTGTAC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCCTCTTTATATGTGGCCCTTTCAAAAGGCCCAATTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTTGATGATTCCTGGAAAAGTATANAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTTGCTGGACGAGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGTGAATGACCCAGATATGTTAGTGATTTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTTACCGCTTAGACGGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAACCGCGAGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGATGGGCTGTATCTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAAGGAAGCTAGGGTTCTATGATGAGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAAGATCATATAAATCCACAGGCGACTGTTTGTGTTCTCAGCTA 1251
RESULT 12
AAD45226
ID AAD45226 standard; DNA; 1272 BP.
XX
AC AAD45226;
```

```
XX 27-DEC-2002 (first entry)
XX DE Human rGAL-12R DNA.
XX KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX KW therapeutic; rGAL-12R; gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX CDS 1..1272
FT /tag= a
FT /product= "Human rGAL-12R protein"
XX PN US2002088024-A1.
XX XX
XX PD 04-JUL-2002.
XX PF 13-NOV-2001; 2001US-00993059.
XX PR 26-JUL-2000; 2000US-00626127.
XX XX
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX XX
XX WPI; 2002-681656/73.
XX DR P-PSDB; AAE28213.
XX XX
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX storage diseases.
XX PS Claim 1; Page 50-51; 88pp; English.
XX CC The invention relates to human alpha-galactosidase truncated at the
XX CC carboxy terminus and the production of enzymatically active recombinant
XX CC human and animal lysosomal enzymes. The invention is useful for producing
XX CC lysosomal enzymes for treating lysosomal storage diseases, producing
XX CC altered or mutated proteins, enzymatically active or otherwise, to serve
XX CC as precursors or substrates for further in vivo or in vitro processing to
XX CC a specialised industrial form for research or therapeutic uses, to
XX CC produce more effective therapeutic enzyme, for producing antibodies
XX CC against lysosomal enzymes for medical diagnostic use, and in any
XX CC commercial process that involves substrate hydrolysis. The present
XX CC sequence is human rGAL-12R DNA
XX SQ Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 2,39e-235 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
```

US-10-602-220-16 (1-417) x AAD45226 (1-1272)

```
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGCCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAAGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACTGGGAGCGGCTTCATGTGCNACCTTGACTGCCAGGAGAGCCA 180
```

```
QY 61 AspSerCysIleSerGluIysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTTCATCATAGTGAAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyIleValGluIysLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGNAGGATGCAGTTATGAGTACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAGGCAGACTTCAGGCAGACCCCTCAGCGCTTTCCTCATGGGATTCGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerIysGlyLeuIysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTTACAGCAAGAGGACTGAAGCTAGGGAATTTATGCAGATGTTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyIleTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCCTGGAGTGTGGATACCTACGACATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysIleAspAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTGTGTACTGTACAGATTTTGGAAAAATTG 540
QY 181 AlaAspGlyTyrIlyshisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATGTGTGTC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCTGTGAGTGGGCTCTTTATATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAAATATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspAlaGlyProGly 260
DB 721 AGTATCTTGACATGGACATCTTTAAACCAGAGAGAAATTTGATGTGCTGGACAGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGNATGACCCAGATATGTTAGTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATTAAGGACGTAATGGCATCAAT 960
QY 321 GlnAspProLeuGlyIysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGACCCCTTGGCAGCAGGAGGTACCACTAGACAGGAGACCACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCCGCGAGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGAACCTCGCTTTATACCATCGCAGTTGCTTCCCTGGGTAAAGAGTGGCCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValIysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGACTCTCTCCCTGTGNAAGGAGAGCTAGGGTCTCATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAGTCACATAAAATCCACAGGCACCTGTTTGTCTTCAGCTA 1251
```

RESULT 13

```
ADD84756
ID ADD84756 standard; DNA; 1272 BP.
XX AC ADD84756;
XX 29-JAN-2004 (first entry)
XX Human alpha-galactosidase rGAL-12R DNA.
XX Human; alpha-galactosidase; rGAL-4; gene; ds; lysosomal enzyme;
XX enzyme replacement therapy; lysosomal disease.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 1..1272
XX FT /*tag= a
XX FT /product= "Human rGAL-12R"
XX PN US2003106095-A1.
XX XX
XX PD 05-JUN-2003.
XX PF 20-MAR-2002; 2002US-00103327.
XX PR 26-JUL-2000; 2000US-00626127.
XX PR 13-NOV-2001; 2001US-00993059.
XX XX
XX PA (GARG/) GARGER S J.
XX PA (TURP/) TURPEN T H.
XX PA (KUMA/) KUMAGAI M H.
XX XX
XX PI Garger SJ, Turpen TH, Kumagai MH;
XX XX
XX DR WPI; 2003-801257/75.
XX DR P-PSDB; ADD84757.
XX PT New polynucleotide for producing active recombinant human and animal
XX PT lysosomal enzymes in a plant expression system that can be used in enzyme
XX PT replacement therapy.
XX PS Claim 1; SEQ ID NO 17; 77pp; English.
XX CC The invention relates to human alpha-galactosidase derivatives and the
XX CC nucleic acids encoding them. The polypeptides are used in a method for
XX CC producing active recombinant human and animal lysosomal enzymes in a
XX CC plant expression system. The enzymes can be used in enzyme replacement
XX CC therapy for the therapeutic treatment of human and animal lysosomal
XX CC diseases. This sequence represents DNA encoding a human alpha-
XX CC galactosidase derivative polypeptide of the invention.
XX SQ Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,39e-235 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-602-220-16 (1-417) x ADD84756 (1-1272)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGCTTCCTCGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTGTTTCTTGGGACATCCCTCGGGCTAGAGCACTGGACANTGGATTGGCAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
```


Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-220-16 (1-417) x ADJ88282 (1-1272)

QY 1 MetGlnLeuArgAsnProGlnLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTGCCTTCCTCGGCC 60

QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTCTCGGACATCCTGGGCTAGAGCACTGCACATGATGTCGCAAGGACGCT 120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGCA 180

QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCCTGCATCATGTCAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAAGATGCAGTTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCCAAGA 300

QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTCAGAAAGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCCTCATGGATTCGCCAGCTA 360

QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATTTATGTTCCACAGCAAGAGCTGAAGCTAGGAGATTTATGCAGATGTGGAATAAA 420

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTCGGAGCTTTGGATCTACGACATTCATGCCAGACCTTTGCT 480

QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAAATTTGATGTTGTTACTGTGACAGCTTTGGAAAAATTG 540

QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTATAGCACATGTCTTGCCCTGAATAGGACTGGCAGAGCATTTGTGTAC 600

QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCAATTTATACAGAAATCCGA 660

QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCAATCACTGGCGAAATTTTGCTGACATTTGATGATTCCTGGAAAAAGTATAAG 720

QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACTGGACATCTTTTAAACCAGGAGAAATTTGATGTTGCTGACACCGGG 780

QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTGGNATGACCAGATAGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGATCAGCAA 840

QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTCATCATGGCTGCTCTTTATTTATGTCCTAATGACCTC 900

QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCCTTCAGGATAGGACGTAATTTGCCATCAAT 960

QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAAGGGTACACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATTAACCGGAGGATGGT 1080

QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGAGTTGCTTCCCTGGGTAAAGGAGTGGCTGTATCCT 1140

QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACAGCTCCTCCTGTGAAAGAAAGCTAGGGTCTTATGAATGACT 1200

QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
DB 1201 TCAAGGTTAAGAAGTCAATATAATCCACAGGCACTGTTTTGCTTCAGCTA 1251

RESULT 15
ADM48686
ID ADM48686 standard; DNA; 1272 BP.
XX
AC ADM48686;
XX
DT 03-JUN-2004 (first entry)
XX
DE Human wild type rGAL-12R DNA.
XX
KW Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW gene; ds.
XX
OS Homo sapiens.
XX

Key Location/Qualifiers
FH 1.1272
FT CDS
FT /*tag= a
FT /product= "GAL-12R protein"
XX
US2004023281-A1.
XX
XX 05-FEB-2004.
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX

(TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
PA (POGU/) POGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRILL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
XX WPI; 2004-142650/14.
XX P-PSDB; ADM48687.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.

XX Disclosure; SEQ ID NO 17; 72pp; English.
XX The present invention relates to novel galactosidase (Gal) proteins such
CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
CC of the present invention are useful for producing recombinant lysosomal
CC enzymes for enzyme replacement therapy for treating human and animal
CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
CC human wild type rGAL-12R DNA. This sequence comprises a human rGAL-12 DNA
CC and ER retention signal DNA. This sequence is used in the exemplification
CC of the invention.

XX SQ Sequence 1272 BP; 323 A; 291 C; 333 G; 325 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,39e-235 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-602-220-16 (1-417) x ADM48686 (1-1272)

Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTGGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTCCTGGGACATCTCCCTGGGGCTAGAGCACTGGACATGGATTGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGGACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCATCAGTGAGAGAGCTCTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
Qy 81 TrpLysAspAlaGlyTrpGluTrpLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAAGGAGCAGACTTCAGGAGAGACCTCAGGCGCTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTrpValHisSerLysGlyLeuLysLeuGlyIleTrpAlaAspValGlyAsnLys 140
Db 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGATTTATGACAGATGTTGGAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTrpTrpAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGGATCTACGACATTCATGATGATGATGATGATGAT 480
Qy 161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTrpCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 540
Qy 181 AlaAspGlyTrpLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTrp 200
Db 541 GCAGATGTTATAGCAGCATGCTCTGGCCCTGAAATAGGACTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTrpProLeuTrpMetTrpProPheGlnLysProAsnTrpThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTrpCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGGTGTGACATTGATGATTCCTGGAAAGTATAAAG 720

Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGGACTGGACATCTTTAAACCAGGAGAGAAATTTTGTATGTTGCTGGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGGATGACCCAGATATGTTAGTGGCACTTTGGCCTCAGCTGGATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTTATTTATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACATCATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAGGGGTACCACTTAGACAGGAGAGCAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTrpGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGTAAGGAGGAGCTAGGCTTCTATGATGAGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTAAAGAGTCAATAAATCCACAGGCACTGTTTGTGCTTCAGCTA 1251

Search completed: January 1, 2006, 00:49:52
Job time : 1146.75 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2006, 00:27:04 ; Search time 9516.23 Seconds
(without alignments)
2490.874 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNPDLHLCALALFLA.....EWSRLRSHINPTGTVLLQL 417

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US10602220/runat_30122005_140557_5421/app_query.fasta_1.782
-DB=GenEmbl -QFWT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10602220@cgn 1.1 5556 @runat_30122005_140557_5421 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database :

GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*

ALIGNMENTS

RESULT 1
AR653698 LOCUS AR653698 1254 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 15 from patent US 6887696.
ACCESSION AR653698
VERSION AR653698.1 GI:67584283
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1254)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 15 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA

FEATURES
source
1..1254
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	1254	6	AR653698 Sequence
2	2293	100.0	1254	6	AR654966 Sequence
3	2293	100.0	1266	6	AR653696 Sequence

Alignment Scores:				
Pred. No.:	1.29e-211	Length:	1254	
Score:	2293.00	Matches:	417	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	

4	2293	100.0	1266	6	AR654964	AR654964 Sequence
5	2293	100.0	1272	6	AR653699	AR653699 Sequence
6	2293	100.0	1272	6	AR654967	AR654967 Sequence
7	2293	100.0	1278	6	AR653694	AR653694 Sequence
8	2293	100.0	1278	6	AR654962	AR654962 Sequence
9	2293	100.0	1284	6	AR653697	AR653697 Sequence
10	2293	100.0	1284	6	AR654965	AR654965 Sequence
11	2293	100.0	1290	6	CS135670	CS135670 Sequence
12	2293	100.0	1290	6	AR653692	AR653692 Sequence
13	2293	100.0	1290	6	AR654960	AR654960 Sequence
14	2293	100.0	1290	8	BT006864	BT006864 Homo sapi
15	2293	100.0	1290	11	AY889631	AY889631 Synthetic
16	2293	100.0	1290	11	AY889632	AY889632 Synthetic
17	2293	100.0	1290	11	AY892090	AY892090 Synthetic
18	2293	100.0	1290	11	BT007835	BT007835 Synthetic
19	2293	100.0	1296	6	AR653695	AR653695 Sequence
20	2293	100.0	1296	6	AR654963	AR654963 Sequence
21	2293	100.0	1306	6	BD182986	BD182986 Pharmaceu
22	2293	100.0	1306	6	BD185988	BD185988 Pharmaceu
23	2293	100.0	1308	6	AR653693	AR653693 Sequence
24	2293	100.0	1308	6	AR654961	AR654961 Sequence
25	2293	100.0	1312	8	BC002689	BC002689 Homo sapi
26	2293	100.0	1319	6	AX840234	AX840234 Sequence
27	2293	100.0	1319	8	HSAGALAR	X05790 Human mRNA
28	2293	100.0	1343	6	AR101984	AR101984 Sequence
29	2293	100.0	1343	6	BD064229	BD064229 Therapy f
30	2293	100.0	1343	6	BD247365	BD247365 Treatment
31	2293	100.0	1343	6	CS105027	CS105027 Sequence
32	2293	100.0	1343	6	AR234439	AR234439 Sequence
33	2293	100.0	1343	6	AR322462	AR322462 Sequence
34	2293	100.0	1343	6	AR372138	AR372138 Sequence
35	2293	100.0	1350	6	CQ899396	CQ899396 Sequence
36	2293	100.0	1350	6	CS033394	CS033394 Sequence
37	2293	100.0	1350	6	CS035758	CS035758 Sequence
38	2293	100.0	1350	6	CS042346	CS042346 Sequence
39	2293	100.0	1350	6	CS044710	CS044710 Sequence
40	2293	100.0	1352	6	AX523899	AX523899 Sequence
41	2293	100.0	1390	6	AX780273	AX780273 Sequence
42	2293	100.0	1393	6	I30313	I30313 Sequence
43	2293	100.0	1393	6	AX030659	AX030659 Sequence
44	2293	100.0	1447	6	BD134457	BD134457 Human nuc
45	2293	100.0	1447	6	AX017286	AX017286 Sequence

DB:	6	Gaps:	0
US-10-602-220-16 (1-417) x AR653698 (1-1254)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGCTGGCTGACCTGGGAGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCTGTCATCAGTGGAGCTTCATGGAGTGGCAGAGCTCATGGTTCAGAGGCC	240
QY	81	TrpLysAspAlaGlyTyrGluTyrLeuGluCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGTGGCTCCCAAGA	300
QY	101	ArgHisIleSerProGlnAlaLeuAlaLeuGlnAspLysAspValIleAlaIleAsn	320
Db	301	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTTAATGGCCATCAAT	360
QY	321	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGGCAAGGATACGAGCTTACAGAGGAGACAACTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluGly	360
Db	1021	GAACGACCTCTCTAGGCTTAGCTGGCTGTAGCTATGATAAACCCGAGGAGATTGGT	1080
QY	361	GlyProArgSerTrpThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
Db	1081	GGACCTCGCTCTTATACCATCGCAGTGTCTCCCTGGGTAAAGAGTGGCTGTAAATCCT	1140
QY	381	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
Db	1141	GCCTGCTTTCATCACACAGCTCTCCCTGTGAAAAAGGAAGCTAGGGTTCTATGAATGGACT	1200
QY	401	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
Db	1201	TCAAGGTAAAGTCACTAATAATCCACAGGCACTGTTTGTCTTCAGCTA	1251
RESULT 2			
LOCUS	AR654966	1254 bp	DNA linear PAT 13-JUN-2005
DEFINITION	Sequence 15 from patent US 6890748.		
ACCESSION	AR654966		
VERSION	AR654966.1	GI:67586505	
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 1254)		
AUTHORS	Gargner, S.J., Turpen, T.H. and Kumagai, M.H.		
TITLE	Production of lysosomal enzymes in plants by transient expression		
JOURNAL	Patent: US 6890748-A 15 10-MAY-2005;		
FEATURES	Large Scale Biology Corporation; Vacaville, CA		
source	Location/Qualifiers		
	1. .1254		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Alignment Scores:	1.29e-211	Length:	1254
Pred. No.:	2293.00	Matches:	417
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	6		
US-10-602-220-16 (1-417) x AR654966 (1-1254)			
QY	1	MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla	20
Db	1	ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTCGGCTTCCTGGCC	60
QY	21	LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
Db	61	CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGGATTGGCAAGACGCT	120
QY	41	ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro	60
Db	121	ACCATGGCTGGCTGACCTGGGAGCTTCATGTGCAACCTTGACTGCCAGGAAGCCCA	180
QY	61	AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly	80
Db	181	GATTCTGTCATCAGTGGAGCTTCATGGAGTGGCAGAGCTCATGGTTCAGAGGCC	240
QY	81	TrpLysAspAlaGlyTyrGluTyrLeuGluCysIleAspAspCysTrpMetAlaProGlnArg	100
Db	241	TGGAAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGTGGCTCCCAAGA	300
QY	101	ArgHisIleSerProGlnAlaLeuAlaLeuGlnAspLysAspValIleAlaIleAsn	320
Db	301	CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAGGAGCTTAATGGCCATCAAT	360
QY	121	GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
Db	961	CAGGACCCCTTGGGCAAGGATACGAGCTTACAGAGGAGACAACTTTGAAGTGTGG	1020
QY	341	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluGly	360

QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACATACGACATTCATGCCAGACCTTGTCT 480
 QY 161 AspTyrGlyValAspIleLeuLeuPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTGTACTGTGACAGTTGGGAAATTTG 540
 QY 181 AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATTAAGACATGCTCTTGGCCCTGATAGGACTGGCAGAACATTTGTGAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGGCCAATATATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpIleSerIleLys 240
 DB 661 CAGTACTGCATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAGATATAAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGACCTGGACATCTTTTACCAGAGAGATTTGTGATTTGTGTCGCCAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGAAATCACCCAGATATGTTAGTGGTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTTCTAATGACCTC 900
 QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 DB 901 CGACACATCAGCCCTCAGGCCAAAGCTCTCTTCAGGATTAAGACGTTAATGGCATCAAT 960
 QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 DB 961 CAGGACCCCTTGGCAAGCAAGGTACACAGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
 QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 DB 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATNAACCGGCAGAGATGGT 1080
 QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 DB 1081 GGACCTCGCTTTATACCATCGCAGTTGCTTCTCTGGGTAAAGAGTGGCTGTATCCT 1140
 QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
 DB 1141 GCCTGCTTCATCACAGACCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGACT 1200
 QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 DB 1201 TCAAGGTTAAGAGTCAATAAAATCCACAGGCACTGTTTTGTCTCAGCTA 1251
 RESULT 3
 AR653696
 LOCUS AR653696 1266 bp DNA linear PAT 13-JUN-2005
 DEFINITION Sequence 11 from patent US 6887696.
 ACCESSION AR653696
 VERSION AR653696.1 GI:67584281
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 1266)
 AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
 TITLE Production of lysosomal enzymes in plants by transient expression
 JOURNAL Patent: US 6887696-A 11 03-MAY-2005;
 FEATURES Large Scale Biology Corporation; Vacaville, CA
 Location/Qualifiers

source

1. 1266

 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Alignment Scores:

 Pred. No.: 13e-211 Length: 1266
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x AR653696 (1-1266)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCGCTTCCTCGGCC 60
 QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 DB 61 CTCGTTTCTCTGGGACATCCCTGGGGCTAGAGCACTGGCAATGGATTTGGCAAGACGCT 120
 QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 DB 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
 QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 DB 181 GATTCCTGCATCAGTGGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAAGC 240
 QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 DB 241 TCGAAGGATGACGGTTATGAGTACCTCTGCAATTGATGACTGTTGGATGGCTCCCAAGA 300
 QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 DB 301 GATTCAGAGGAGAGACTTCAGGAGACCCCTCAGCGCTTCTCATGGATTCGCAGCTA 360
 QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 DB 361 GCTAATTTATGTTACAGCAAAAGGACTGAAGCTAGGAGTTATGCAGATTTGGAAATAA 420
 QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 DB 421 ACCTGCGCAGGCTTCCCTGGGAGTTTGGATACATACGACATTCATGCCAGACCTTTGCT 480
 QY 161 AspTyrGlyValAspIleLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 DB 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTTGTACTGTGACAGTTTGGAAATTTG 540
 QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 DB 541 GCAGATGGTTATTAAGCACATGCTCTTGGCCCTGATAGGACTGGCAGAGCATTTGTGAC 600
 QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 DB 601 TCCTGTGAGTGGCTCTTTATATATGTGGCCCTTTCAAAAGGCCAATATATACAGAAATCCGA 660
 QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpIleSerIleLys 240
 DB 661 CAGTACTGCATCACTGGCGAAATTTTGTGACATTCATGATTCCTGGAAAGATATAAAG 720
 QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 DB 721 AGTATCTTGACCTGGACATCTTTTACCAGAGAGAAATTTGTGATTTGTGTCGCCAGGG 780
 QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
 DB 781 GGTGGAAATCACCCAGATATGTTAGTGGTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
 QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300

```
Db      841  GTAACATCAGATGSCCCTCTGGGCTATCATGGCTGCTCCTTTATTATCATGCTTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCAAGCTCTCTTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGAGCAACTTTGAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGCTGTAGCTATGATAAACCGCAGGAGATTGCT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      1141  GCCTGCTTCATCACACAGCTTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201  TCAAGCTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 4
AR654964
LOCUS      1266 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION      Sequence 11 from patent US 6890748.
ACCESSION      AR654964
VERSION      AR654964.1  GI:67586503
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      Unclassified.
AUTHORS      Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE      Production of lysosomal enzymes in plants by transient expression
JOURNAL      Patent: US 6890748-A 11 10-MAY-2005;
              Large Scale Biology Corporation; Vacaville, CA
FEATURES
source      1..1266
              /organism="unknown"
              /mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.:      1.3e-211      Length:      1266
Score:      2233.00      Matches:      417
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      6      Gaps:      0

US-10-602-220-16 (1-417) x AR654964 (1-1266)
Qy      1  MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db      1  ATCGAGCTGAGGAACCCAGAACTACATCTGGGCTGCGGCTTGGCTTCTCGCTTCTTGGCC 60
Qy      21  LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db      61  CTCGTTTCTCTGGGACATCCCTGGGGCTTAGAGCACTGGACAATGGATTGGCAAGAGCCCT 120
Qy      41  ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db      121  ACCATGGGCTGGCTGCATCTGGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy      61  AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db      181  GATTCTCTGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAGGC 240
```

```
Qy      81  TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db      241  TGGAGGATGCAAGTTATGAGTACCTCTGCATTGATGATGCTGTGGATGCTCCCAAGA 300
Qy      101  AspSerGluGlyArgLeuGlnAlaAspProGlnAlaArgPheProHisIleArgGlnLeu 120
Db      301  GATTCAAGAGGAGAGCTTCAGGCAGACCCCTCAGCGCTTTCCTCATCGGATTCGCCAGCTA 360
Qy      121  AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAlaAspValGlyAsnLys 140
Db      361  GCTAATATATGTTTACAGCAAGAGACTGAAGCTAGGATTTATGCAGATGTTGGAAATAA 420
Qy      141  ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db      421  ACCTGGCAGGCTTCCCTGGGAGTTTGTGATACTACGACATTGATGCCAGACCTTTGCT 480
Qy      161  AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db      481  GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540
Qy      181  AlaAspGlyTyrIleHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db      541  CGAGATGGTTATAAGCACATGCTCTTGGCCCTGAAATAGGACTGGCAGAGCATTTGTATC 600
Qy      201  SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db      601  TCCTGTGAGTGGCTCTTTATATGTGGCTTTTCAAAAGCCCAATTTATACAGAAATCCGA 660
Qy      221  GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db      661  CAGTACTGCAATCAGCTGGGGAATTTTGTGACATTGATGATTCCTGGAAAAAGTATAAAG 720
Qy      241  SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db      721  AGTATCTTGGACTGGACATCTTTTAAACGAGAGAAATTTGTTGATGTTGCTGACACAGG 780
Qy      261  GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db      781  GGTGTGAATGACCCAGATATGTTAGTGTGGCACTTTGGCCTCAGCTGGAATCAGCAA 840
Qy      281  ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db      841  GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCCTTTATTCATGCTTAATGACCTC 900
Qy      301  ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db      901  CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTCGCATCAAT 960
Qy      321  GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db      961  CAGGACCCCTTGGGCAAGCAAGGTAACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
Qy      341  GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db      1021  GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT 1080
Qy      361  GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db      1081  GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGAGTGGCTGTATCCT 1140
Qy      381  AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db      1141  GCCTGCTTCATCACACAGCTTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
Qy      401  SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db      1201  TCAAGCTTAAGAGTCACATAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 5
AR653699
LOCUS      1272 bp      DNA      linear      PAT 13-JUN-2005
DEFINITION      Sequence 17 from patent US 6887696.
ACCESSION      AR653699
```

```
VERSION AR653699.1 GI:67584284
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 17 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1272
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,31e-211 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR653699 (1-1272)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTCACACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpIleAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACACAGCAAGGACTGAAGCTAGGATTTATGACATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACATGTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACACATGCTCTGGCCCTGAAATAGGACTGGCAGAACATTTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGGCTGACATTTGATGATTTCTTGGAAAGTATAAG 720
```

```
Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGATTTGTATGTTGTGGACCAAGG 780
Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGGATACCCAGATATGTTAGTATGGCACTTTGGCCCTCAGCTGGATCAGCA 840
Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTCGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGTAATTTGCCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGGCAAGCAAGGTCACGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGACCTCTCTCAGCCTTAGCTGGCTGATGATGATAACCGCAGGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGCACCTCGCTCTTATACCATCGAGTTGCTTCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValIysArgIleGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCCCTGTGAAAGGAAGCTAGGCTTCTATGAATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAGGTTAAGAAGTACATATAATCCACAGGACCTGTTTGTCTTCAGCTA 1251
RESULT 6
LOCUS AR654967 1272 bp DNA linear PAT 13-JUN-2005
DEFINITION Sequence 17 from patent US 6890748.
ACCESSION AR654967
VERSION AR654967.1 GI:675866506
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Garger, S.J., Turpen, T.H. and Kumagai, M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6890748-A 17 10-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1272
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,31e-211 Length: 1272
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR654967 (1-1272)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTCCTCGCC 60
Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCCTGGGACATCCCTGGGCTAGAGCACTGGCAATGATGGCAAGACGCT 120
Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGGCTGGCTGGCACTGGAGCGCTTCATGTCACACCTTGACTGCCAGGAGGCCA 180
Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATCAGTGAGAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpIleAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAAAGATGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGAGACTTCAGGAGAGCCCTCAGCGCTTTCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATGTTACACAGCAAGGACTGAAGCTAGGATTTATGACATGTTGGAAATAAA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGCTTCCTGGGAGTTTGGATACCTACGACATGATGCCAGACCTTTGCT 480
Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGACATGTTGGAAATTTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGACACATGCTCTGGCCCTGAAATAGGACTGGCAGAACATTTGTAC 600
Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATATGTCCTTTTCAAAAGCCCAATTATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGGCTGACATTTGATGATTTCTTGGAAAGTATAAG 720
```

Db 61 CTGCTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGATGGCAAGAGCGCTT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTCCACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGTACTAGCAATGATGTCGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTACTGTGACGATTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATTAAGCACATGCTCTGGCCCTGAATAGGACTGGCAGAAAGCATTTGTATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCTGTGAGTGGCTCTTTATATATGTGGCCCTTCAAAGGCCCAATTAACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACTGGCGAAATTTTGTGTCATGATGATGATGATGATGATGATGATGAT 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnLysIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACTGGACATCTTTTAAACAGGAGAGAAATGTTGATGTTGCTGGACCGGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
Db 781 GGTGGNATGACCCAGATATGTTAGTGTGGCACTTTGGCCCTCAGCTGGATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGCTTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCAAGCTCTCTTCAGGATAAGACGCTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCAGCAGGAGTACAGCTTAGACAGGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAAGGACCTCTCTCAGGCTTAGCGCTGTAGCTATGATAAACCGCAGGAGATTGCT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGAACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGAGTGGCTGTAACTCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400

Db 1141 GCCTGCTTCATCACAGACTCCTCCCTGTGTAAGGAAGAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAAGTACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251
RESULT 7
AR653694 1278 bp DNA linear PAT 13-JUN-2005
LOCUS Sequence 7 from patent US 6887696.
DEFINITION AR653694
ACCESSION AR653694
VERSION AR653694.1 GI:67584279
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1278)
AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
TITLE Production of lysosomal enzymes in plants by transient expression
JOURNAL Patent: US 6887696-A 7 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1278
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,32e-211 Length: 1278
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR653694 (1-1278)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAAACCCAGAACTACATCTGGGCTGGCGCTTGGCTTGGCTTCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTCGTTTCTGGGACATCCCTGGGGCTAGAGCACTGGACATGGATGGCAAGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCTCGCATAGTCAGGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGGAGGATGCAGGTATAGTACCTCTGCATTGATGACTGTGGATGGCTCCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAAGGCAGACTTCAGGCAGACCCCTCAGCGCTTCTCATGGATTGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATATATGTTCCACAGCAAGAGCTGAAGCTAGGATTTATGCAGATGTGGAAATAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGCGAGGCTTCCCTGGGAGTTTGGTACTAGCAATGATGTCGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGGAGTAGACTCTGCTAAATTTGATGGTTGTACTGTGACGATTTGGAAATTTG 540

Qy 181 AlaaspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGACACATGCTCCCTGGCCCTGAAATAGGACTGGCAGAGCAATGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCCGTGAGTGGCCCTCTTTATATATGCGCCCTTTCAAAGGCCAATATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGCGCAATTTTGTGTGACATGATGATCTCTGAAAAGTATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATGTTGCTGGACCAAGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGCTATCATGCTGCTCTCTTTATTCATGTCTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGCAAGCAAGGGTACAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020
 Qy 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
 Db 1021 GAACGACTCTCTCAGCTTAGCTGGCTGGCTGTAGTATGATTAACCGCAGGAGATGGT 1080
 Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
 Db 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCCTGGGTAAGGAGTGGCTGTAAATCCT 1140
 Qy 381 AlaCysPheIleThrGlnLeuLeuProValIleArgLysLeuGlyPheTyrGluTrpThr 400
 Db 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAGGCTAGGGTTCTATGATGAGACT 1200
 Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
 Db 1201 TCAAGGTTAGAGTCACTAATAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 8
 AR654962
 LOCUS 1278 bp DNA linear PAT 13-JUN-2005
 DEFINITION Sequence 7 from patent US 6890748.
 ACCESSION AR654962
 VERSION AR654962.1 GI:67586501
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1278)
 AUTHORS Garger,S.J., Turpen,T.H. and Kumagai,M.H.
 TITLE Production of lysosomal enzymes in plants by transient expression
 JOURNAL Patent: US 6890748-A 7 10-MAY-2005;
 Large Scale Biology Corporation; Vacaville, CA
 FEATURES
 source Location/Qualifiers
 1..1278
 /organism="unknown"
 /mol_type="genomic DNA"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.32e-211 Length: 1278
 Score: 2293.00 Matches: 417
 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-10-602-220-16 (1-417) x AR654962 (1-1278)
 Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
 Db 1 ATCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCGCTTGGCTTCGCTTCCTCGGCC 60
 Qy 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
 Db 61 CTGTTTCTCTGGGACATCCCTGGGCTAGAGCACTGGACAAATGATTTGGCAGACGCGCT 120
 Qy 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
 Db 121 ACCATGGGCTGGCTGCACCTGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCA 180
 Qy 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
 Db 181 GATTCTCTGCATCAGTGAGAAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTTCAGAAAGC 240
 Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
 Db 241 TGGAGATGTCAGGTTATGATGATCTCTGATTTGATGATCTTTGGATGGCTCCCAAGA 300
 Qy 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
 Db 301 GATTCAAGAGGACAGACTTCAGGCAGACCCCTCAGCGCTTTCTCATGGATTGGCAGCTA 360
 Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
 Db 361 GCTAATATATCTTCAACGCAAGGACTGAAGCTAGGGATTTATGCAGATGTTGGAAATAA 420
 Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
 Db 421 ACCTGGCAGGCTTCCCTGGGAGTTTGGATACCTACGACATGATGATGCCAGACCTTTGCT 480
 Qy 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
 Db 481 GACTGGGAGTAGACTGCTGCTAATAATTGATGGTTGTACTGTGACAGTTTGGAAATTTG 540
 Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
 Db 541 GCAGATGGTTATAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCAATGTGTAC 600
 Qy 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
 Db 601 TCTGTGAGTGGCTCTTTATATGTTGGCCCTTTCAAAGGCCAATATACAGAAATCCGA 660
 Qy 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspSerTrpLysSerIleLys 240
 Db 661 CAGTACTGCAATCACTGGCGAATTTTGTGTGACATGATGATCTCTGAAAAGTATAAG 720
 Qy 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
 Db 721 AGTATCTTGGACTGGACATCTTTTAAACGAGGAGAAATTTGTGATGTTGCTGGACAGGG 780
 Qy 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln 280
 Db 781 GGTGGAAATGACCCAGATATGTTAGTATGTTGGCACTTTGGCCCTCAGCTGGAAATCAGCAA 840
 Qy 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
 Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGCTGCTCTTTATTCATGTCTAATGACCTC 900
 Qy 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
 Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCCAGGATAAGGACGTAATTTGCCATCAAT 960
 Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
 Db 961 CAGGACCCCTTGGCAAGCAAGGGTACAGCTTAGACAGGAGACAATTTGAAGTGTGG 1020

QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGCTTAGCTGGCTGTAGCTATGATATAACCGCGAGGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTATACCATCGCAGTGTCTTCCCTGGGTAAAGAGTGGCTGTATCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuLeuLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACGTGTTTGTCTCAGCTA 1251

RESULT 9
AR653697 1284 bp DNA linear PAT 13-JUN-2005
LOCUS
DEFINITION Sequence 13 from patent US 6887696.
ACCESSION AR653697
VERSION AR653697.1 GI:67584282
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 1284)
TITLE Garger, S.J., Turpen, T.H. and Kumagai, M.H.
JOURNAL Production of lysosomal enzymes in plants by transient expression
Patent: US 6887696-A 13 03-MAY-2005;
Large Scale Biology Corporation; Vacaville, CA
FEATURES
source Location/Qualifiers
1..1284
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Alignment Scores:
Pred. No.: 1,33e-211 Length: 1284
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-602-220-16 (1-417) x AR653697 (1-1284)

QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAACCTACATCTGGGCTGCGCGCTTGGCTTCTCGTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTCTCGGACATCCCTGGGCTTAGAGCACTGGCAATGGATTGGCAAGGACGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
Db 121 ACCATGGGCTGGGCTGACCTGGAGCGCTTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTTCCTGCAATCAGTGAGAGCTCTTCATGGAGTGGCAGAGCTCATGGTCTCAGAGGC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
Db 241 TGAAGGATGCAGGTATATGATACCTCTGCATTGATGACTGTGGATGGCTCCCAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTTCAGAGGCAGACTTCAGGCAGACCTCAGCGCTTCTCATGGATTGGCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140

Db 361 GCTAATATATGTTTACAGCAAGGACTAGGAATTTATGCAAGTTGTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTTGGATACCTACGACATGATGATGCCAGACCTTGGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAAATTTGATGGTTGTTACTGTGACAGTTTGGAAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAACATTTGTGATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCCTCTTTATATGTGGCCCTTTCAAAGCCCAATTTATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
Db 661 CAGTACTGCAATCACCTGGCGAAATTTTGTCTGACATTTGATGATTCCTGGAAAAGTATAAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTATCTTGACCTGGACATCTTTTAAACAGGAGAGAAATTTGTGATGTGTGGACCAGGG 780
QY 261 GlyTrpAsnAspProAsnMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
Db 781 GGTGTGAATGACCCAGATATGTTAGTGATTTGGCACTTTGGCCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
Db 841 GTAACCTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCTAATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
Db 901 CGACACATCAGCCCTCAAGCCAAAGCTCTCTTCAGGATAAGGACGCTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
Db 961 CAGGACCCCTTGGCGAAGCAAGGTTACCGCTTAGACAGGAGACAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCTGGGCTGTAGCTATGATAAACCGCAGGAGATGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGCAGTGTGCTTCCCTGGGTAAAGAGTGGCTGTAAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
Db 1141 GCCTGCTTCATCACAGCTCTCTCCCTGTGAAAGGAAGCTAGGGTTCTATGAATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuLeuLeu 417
Db 1201 TCAAGGTTAAGAAGTCACATAAATCCACAGGCACGTGTTTGTCTCAGCTA 1251

RESULT 10
AR654965 1284 bp DNA linear PAT 13-JUN-2005
LOCUS
DEFINITION Sequence 13 from patent US 6890748.
ACCESSION AR654965
VERSION AR654965.1 GI:67586504
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 1284)
TITLE Garger, S.J., Turpen, T.H. and Kumagai, M.H.
JOURNAL Production of lysosomal enzymes in plants by transient expression
Patent: US 6890748-A 13 10-MAY-2005;

	Qy	MetGlnLeuArgAsnProGluLeuHisIeuGlyCysAlaIeuAlaLeuArgPheLeuAla	20
	Db	ATGCAGCTGAGAACCCAGAACTACATCTGGCGTGC	60
	Qy	LeuValSerTrpAspileProGlyValaArgAlaLeuAspAsnGlyLeuAlaArgThrPro	40
	Db	CTCTGTTCCTGGGAATCCCTGGGGCTAGAGCACTGGCAATGATTGGCAAGACGCCCT	120

QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGGAGCCCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAGGCC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAGGTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAGGACAGCTTCAGGAGACCCCTCAGCGCTTCCTCATGGGATGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTTACACAGCAAGGACTGAAGCTAGGANTTATGCAGATGTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGGAGTTTTGGATACACGACATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGCAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600
QY 201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysProAsnTyrThrGluIleArg 220
DB 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAGGCCCAATATATACAGAAATCCGA 660
QY 221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
DB 661 CAGTACTGCNATCCTGGCGGAAATTTTGTGACATTTGATGATTCCTGGGAAAGTATTAAG 720
QY 241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
DB 721 AGTATCTTGACTGGACATCTTTTAAACCAGGAGAGAAATTTGATGTTGCTGGCAGCGG 780
QY 261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
DB 781 GGTTGGAAATGACCCAGATATGTTAGTGGCAACTTTGGCCCTCAGCTGGAAATCAGCAA 840
QY 281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
DB 841 GTAACTCAGATGGCCCTCTGGGCTATCATGGCTGCTCTTTATTCATGTCATATGACCTC 900
QY 301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
DB 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTTCAGGATAAGACGCTAATTTGCCATCAAT 960
QY 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTrp 340
DB 961 CAGGACCCCTTGGCAGCAAGGATACCACTTAGACAGGAGCAACTTTGAAGTGTGG 1020
QY 341 GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly 360
DB 1021 GAACGACCTCTCAGGCTTAGCTGGGCTGTAGCTATGATATAACCGGACAGGAGATTGGT 1080
QY 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
DB 1081 GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTGGGTAAGAGTGGCTCTGTAATCCT 1140
QY 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr 400
DB 1141 GCCTGCTTCATCACACAGCTCCTCCCTGTGAAAGGAAGCTAGGGTTCTATGATGGACT 1200
QY 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417

DB 1201 TCAGGTTAAGAAGTCACATAAATCCACAGGCACCTGTTTGTCTCAGCTA 1251
RESULT 12
AR653692
LOCUS
DEFINITION
AR653692
VERSION
AR653692.1 GI:67584277
KEYWORDS
SOURCE
ORGANISM
Unkown.
Unkown.
Unclassified.
REFERENCE
1 (bases 1 to 1290)
Garger, S.J., Turpen, T.H. and Kumagai, M.H.
AUTHORS
Production of lysosomal enzymes in plants by transient expression
TITLE
Patent: US 6887696-A 3 03-MAY-2005;
JOURNAL
Large Scale Biology Corporation; Vacaville, CA
FEATURES
Location/Qualifiers
1..1290
Source
/organism="unkown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 1,34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-602-220-16 (1-417) x AR653692 (1-1290)
QY 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaIleuAlaLeuArgPheLeuAla 20
DB 1 ATGCAGCTGAGGAACCCAGAACTACATCTGGGCTGCGCTTTGCGCTTCGCTTCCTGGCC 60
QY 21 LeuValSerTrpAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
DB 61 CTCGTTTCTGGGACATCCTCGGGCTAGAGCACTGGACATGGANTGGCAGGAGCGCT 120
QY 41 ThrMetGlyTrpLeuHisTrpGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
DB 121 ACCATGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGGCCA 180
QY 61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
DB 181 GATTCTGCATCAGTCAGAGGCTCTTCATGGAGATGGCAGAGCTCATGTGCTCAGAGGCC 240
QY 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
DB 241 TGGAGGATGCAGGTATGAGTACCTCTGCATTTGATGACTGTGGATGGCTCCCAAAGA 300
QY 101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
DB 301 GATTTCAGAGGACAGCTTCAGGAGACCCCTCAGCGCTTCCTCATGGGATGCCAGCTA 360
QY 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
DB 361 GCTAATATGTTTACACAGCAAGGACTGAAGCTAGGANTTATGCAGATGTGGAAATAAA 420
QY 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
DB 421 ACCTGGCAGGCTTCCTGGGAGTTTTGGATACACGACATTCATGCCAGACCTTTGCT 480
QY 161 AspTrpGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
DB 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTTACTGTGCAGCTTTGGAAATTTG 540
QY 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
DB 541 GCAGATGGTTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCATTTGTATC 600


```

Db      1021 GAACGACCTCTCTCAGGCTTAGCGCTGAGCTATGATAAAACCGCAGCAGAGATTGCT 1080

Qy      361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyIysGlyValAlaCysAsnPro 380
|||||
Db      1081 GGACCTCGCTCTATACCATCGAGTGTCTCCCTGGGTAAGAGGAGTGGCCCTGTAACTCT 1140

Qy      381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
|||||
Db      1141 GCCTGCTTCTATCACAGCTCTCCCTCTGTAAGAGGAGACTAGGTTCTATGATGGACT 1200

Qy      401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
|||||
Db      1201 TCAAGGTTAAGAGTCACATAATCCACAGGCACTGTTTGTCTCAGCTA 1251

RESULT 14
BT006864      1290 bp      mRNA      linear      PRI 13-MAY-2003
LOCUS      Homo sapiens galactosidase, alpha mRNA, complete cds.
DEFINITION      BT006864
ACCESSION      BT006864.1 GI:30582566
VERSION      FLI_CDNA.
KEYWORDS      Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homnidae; Homo.
REFERENCE      1 (bases 1 to 1290)
AUTHORS      Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
      Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
      Phelan,M. and Farmer,A.
TITLE      Cloning of human full-length CDSs in BD Creator(TM) System Donor
      vector
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 1290)
AUTHORS      Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
      Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
      Phelan,M. and Farmer,A.
TITLE      Direct Submission
JOURNAL      Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
      Circle, Palo Alto, CA 94303, USA
COMMENT      This CDS clone is a part of a collection of human full length
      expression clones generated by BD Biosciences Clontech and the
      Harvard Institute of Proteomics. Each CDS has been cloned in two
      forms: with and without stop-codon (to allow fusion with C-terminal
      tag). The CDS has been directionally cloned using BD In-Fusion(TM)
      cloning system between the SalI and HindIII sites of the pDNR-DUAL
      vector. Additional sequences in the clone: 'ACC' after SalI site
      and before 'ATG' to provide Kozak consensus sequence; 'GG' after
      last codon and before HindIII site to maintain reading frame.
      Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES
      Location/Qualifiers
          1..1290
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone_lib="BD Creator(TM) CDS Library derived from MGC
              collection"
              /lab_host="DH5alpha T1 resistant"
              /note="Vector: pDNR-Dual"
          1..1290
              /codon_start=1
              /product="galactosidase, alpha"
              /protein_id="AAP35510.1"
              /db_xref="GI:30582567"
              /translation="MOLRNPELHGCALRFLALVSWDIPGARALDNGLRAPTPTMGW
              LWERPMCNLDQEEPCDISEKLFMEAEALMWSECKDAGVEYLICIDCWMAPOKDS
              EGRLDQPPRFGIRQLANYHSGKLGIYADVGNKTCAGFPSPGFGYDIDAQTFE
              DWGVDLKFPGCYGDSLENADYGHKMSLALNKTGRISIVISCEWPLMPPFKPNYTE
              WNOQVQTMALWAIAPLMSNDLRHSIPQAKALQDKVIADINPDPLGKQGYQLRGLS
              DNFVWERPLSLAWAVAMINROEIGGPRSYTTAVASLGKGVACNPACFTITQLLPVKR
              KLGFEYWTSLRSHINPTGTVLLQLENTMQMSLKDL"
CDS

```

```

ORIGIN
Alignment Scores:      1.34e-211      Length:      1290
Pred. No.:      2293.00      Matches:      417
Score:      100.00%      Conservative: 0
Percent Similarity:      100.00%      Mismatches: 0
Best Local Similarity:      100.00%      Indels:      0
Query Match:      100.00%      Gaps:      0
DB:
US-10-602-220-16 (1-417) x BT006864 (1-1290)
Qy      1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
|||||
Db      1 ATGCAGCTGAGGAACCCAGAACTACACTCTGGGCTGCGCTTGGCGCTTCGCTTCTCTGGCC 60

Qy      21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
|||||
Db      61 CTCGTTCCTCGGACATCCCTGGGCTAGAGCACTGGACAATGGATTGGCAGAGCAGCCT 120

Qy      41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluPro 60
|||||
Db      121 ACCATGGGCTGGCTGCACCTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAAGAGCCA 180

Qy      61 AspSerCysIleSerGluLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
|||||
Db      181 GATTCTCGCATCAGTGAGAAGCTCTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240

Qy      81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTrpMetAlaProGlnArg 100
|||||
Db      241 TGGAGGATGCAGGTTATCAGTACCTCTGCAATTGATGACTCTGTTGGATGGCTCCCAAGA 300

Qy      101 AspSerGluGlyArgLeuGlnAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
|||||
Db      301 GATTTCAGAGGAGCAGACTTCAGGAGAGCCCTCAGCGCTTCTCTCATGGATTCCCAAGCTA 360

Qy      121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
|||||
Db      361 GCTAATTATGTTACACAGCAAGGACTGAAGCTAGGAGTATTATGCAGATGTTGGAATAAA 420

Qy      141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspIleAspIleThrPheAla 160
|||||
Db      421 ACCTGCGCAGGCTCTCCCTGGGAGTTTTGGATATCTACGACATTGATGCCAGACCTTTGCT 480

Qy      161 AspTrpGlyValAspLeuLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
|||||
Db      481 GACTGGGAGTAGATCTGCTAAATTGATGGTTTACTGTGACAGTCTGGAAAATTTC 540

Qy      181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
|||||
Db      541 GCAGATGGTTATAAGCACATGCTCTTGGCCCTGAATAGGACTGGCAGAGCACTTGTGTAC 600

Qy      201 SerCysGluTrpProLeuTyrMetTrpProPheGlnLysPheAsnTyrThrGluIleArg 220
|||||
Db      601 TCCTGTGAGTGGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATTATACAGAAATCCGA 660

Qy      221 GlnTyrCysAsnHisTrpArgAsnPheAlaAspIleAspAspSerTrpLysSerIleLys 240
|||||
Db      661 CAGTACTGCAATCACCTGGCGAAATTTTGTGCAATTGATGATTCCTGGAAAAGTATTAAG 720

Qy      241 SerIleLeuAspTrpThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
|||||
Db      721 AGTATCTTGGACTGGACATCTTTTAACCAAGGAGAGAATTTGTTGATGTTGTCGACCAAGG 780

Qy      261 GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGlnGln 280
|||||
Db      781 GGTTCGAATGACCCAGATATGTTAGTGTGTCGCAACTTTGGCCCTCAGCTCGAATACGCAA 840

Qy      281 ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu 300
|||||
Db      841 GTAACCTCAGATGGCCCTCTGGGGCTATCATGGCTTGCTCTTTATTCATGCTTAATGACCTC 900

Qy      301 ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn 320
|||||

```

```
Db 901 CGACATCAGCCCTCAAGCCAAAGCTCTCTTCAAGATAAGGACGTAATTCATCAAT 960
Qy 321 GlnAspProLeuGlyLysGlnGlyTyrGlnLeuArgGlnGlyAspAsnPheGluValTyr 340
Db 961 CAGGACCCCTTGGCAAGCAAGGTACACAGCTTAGACAGGAGACCACTTTGAAGTGTGG 1020
Qy 341 GluArgProLeuSerGlyLeuAlaTyrAlaValAlaMetIleAsnArgGlnGluIleGly 360
Db 1021 GAACGACCTCTCTCAGGCTTAGCGCTGGCTGTAGCTATGATGATAAACCGGACGAGATTGGT 1080
Qy 361 GlyProArgSerTyrThrIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro 380
Db 1081 GGACCTCGCTCTTATACCATCGAGTTCCTTCCGTGGGTAAAGAGTGGCTGTAAATCCT 1140
Qy 381 AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTyrThr 400
Db 1141 GCCTGCTTCATCACAGACCTCTCCCTGTGTAAGAGGAGCTAGGTTCTATGATGGACT 1200
Qy 401 SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu 417
Db 1201 TCAAGGTTAAGAGTCACATAAAATCCACAGGCACTGTTTGTCTTCAGCTA 1251

RESULT 15
AY889631
LOCUS
DEFINITION
  Synthetic construct Homo sapiens clone FLH025769.01X galactosidase
  alpha (GLA) mRNA, complete cds.
ACCESSION
  AY889631.1 GI:60655716
VERSION
  Human ORF Project.
KEYWORDS
  synthetic construct
SOURCE
  other sequences; artificial sequences.
REFERENCE
  1 (bases 1 to 1290)
  Hines,L., Rolfs,A., Jenson,D., Moreira,D., Raphael,J., Kelley,F.,
  Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
  Williamson,J. and Labaer,J.
  Cloning of human full-length CDS in Creator (TM) recombinational
  vector system
  Unpublished
JOURNAL
  2 (bases 1 to 1290)
  Hines,L., Rolfs,A., Jenson,D., Moreira,D., Raphael,J., Kelley,F.,
  Shen,B., Halleck,A., Koundinya,M., Hu,Y., Zuo,D., Taycher,E.,
  Williamson,J. and Labaer,J.
  Direct Submission
  Submitted (04-JAN-2005) Biological Chemistry and Molecular
  Pharmacology, Harvard Institute of Proteomics, 320 Charles St.,
  Cambridge, MA 02141, USA
COMMENT
  This CDS clone is a part of a collection of human full-length
  expression clones generated by Harvard Institute of Proteomics.
  This ORF clone has been cloned with normalized stop-codon. The CDS
  has been directionally cloned using BD In-Fusion(TM) cloning system
  between the SalI and HindIII sites of the pDR-Dual vector.
  Additional sequences in the clone: 'ACC' after SalI site and
  before 'ATG' to provide Kozak consensus sequence. Each clone is
  clonally isolated and full-length sequence-verified.
  Location/Qualifiers
  1..1290
  /organism="synthetic construct"
  /mol_type="mRNA"
  /db_xref="taxon:32630"
  /clone="FLH025769.01X"
  /lab_host="Escherichia coli DH5alpha T1 resistant"
  /note="derived from MGC template"
  1..1290
  /gene="GLA"
  1..1290
  /codon_start=1
  /genes="GLA"
  /product="galactosidase alpha"
  /protein_id="AA32422.1"
```

```
/db_xref="GI:60655717"
/translation="MQLRNPELHLCALALFIALVSWDIPGARALDNLGARTPTMGW
LHRFMNCLDQEEEDSCISEKLFMEALMYSEGKDGAYEYLCTDDCMAPQRDS
EGRLQADPQRFPHGIRQLANYVHSGIKLGIYADVGNKTCAGFPFGYDYDIDAQTEA
DWGVDLLKDFGVCDSLEADGYSKMSIALNRTGRSIVYSCWPLMYMFPQKPNYTE
IRQYCNHNFNFADIDDSKSIKSLDWTSPNOERIYDVAGPGGNPDMLVIGNFGLS
NQOQVTCMALMAIWAAPLFMSNDLHHSIQAKALLQDKDVIAINQPLGKQGVLRQS
DNFVWRPUSGLAMAVMINRQIEIGGPRSYTIAVASLGGVACNPAFCFITQLLPVXR
KLGFEBWTSRLRSHINPTGTVLLENTQMSLKDLL"

ORIGIN
Alignment Scores:
Pred. No.: 1,34e-211 Length: 1290
Score: 2293.00 Matches: 417
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0

US-10-602-220-16 (1-417) x AY889631 (1-1290)
Qy 1 MetGlnLeuArgAsnProGluLeuHisLeuGlyCysAlaLeuAlaLeuArgPheLeuAla 20
Db 1 ATGCAGCTGAGGAACCCAGAGACTACATCTGGGCTGGCGCTTGCGCTTCGCTTCCCTGCC 60
Qy 21 LeuValSerTyrAspIleProGlyAlaArgAlaLeuAspAsnGlyLeuAlaArgThrPro 40
Db 61 CTGCTTTCCTGGGACATCCCTGGGCTAGAGCACTGGACAATGGATGGCAAGACGCT 120
Qy 41 ThrMetGlyTyrLeuHisTyrGluArgPheMetCysAsnLeuAspCysGlnGluGluPro 60
Db 121 ACCATGGCTGGCTGGCACTGGGAGCGCTTCATGTGCAACCTTGACTGCCAGGAGAGCCA 180
Qy 61 AspSerCysIleSerGlyLysLeuPheMetGluMetAlaGluLeuMetValSerGluGly 80
Db 181 GATTCCTGCATCAGTGAGAGCTTTCATGGAGATGGCAGAGCTCATGGTCTCAGAAGGC 240
Qy 81 TrpLysAspAlaGlyTyrGluTyrLeuCysIleAspAspCysTyrMetAlaProGlnArg 100
Db 241 TGGAGGATCGAGGTATTGATGATACCTCTGCATTTGATGACTGTTGGATGGCTCCCAAGA 300
Qy 101 AspSerGluGlyArgGluAlaAspProGlnArgPheProHisGlyIleArgGlnLeu 120
Db 301 GATTCAGAAGGAGAGCTTCAGGAGAGCCCTCAGCGCTTCTCCTCATGGGATTCGCCAGCTA 360
Qy 121 AlaAsnTyrValHisSerLysGlyLeuLysLeuGlyIleTyrAlaAspValGlyAsnLys 140
Db 361 GCTAATTATGTTACAGCAAGAGGACTGAAGCTAGGATTTATGCAGATGTTGAAATATA 420
Qy 141 ThrCysAlaGlyPheProGlySerPheGlyTyrTyrAspIleAspAlaGlnThrPheAla 160
Db 421 ACCTGGCGAGGCTTCCCTGGGAGTTTGGATACCTAGCAGCATTCATGCCCCAGACCTTTGCT 480
Qy 161 AspTyrGlyValAspLeuLysPheAspGlyCysTyrCysAspSerLeuGluAsnLeu 180
Db 481 GACTGGGAGTAGATCTGCTAAATTTGATGGTTGTACTGTGACAGTTTGGAAAAATTG 540
Qy 181 AlaAspGlyTyrLysHisMetSerLeuAlaLeuAsnArgThrGlyArgSerIleValTyr 200
Db 541 GCAGATGGTTATAGCACATGCTTGGCCCTGATAGAGCTGGCAGAGCATTTGTGTAC 600
Qy 201 SerCysGluTyrProLeuTyrMetTyrProPheGlnLysProAsnTyrThrGluIleArg 220
Db 601 TCCTGTGAGTGGCTCTTTATATGTGGCCCTTTCAAAAGCCCAATATATACAGAAATCCGA 660
Qy 221 GlnTyrCysAsnHisTyrPheAsnPheAlaAspIleAspSerTyrLysSerIleLys 240
Db 661 CAGTACTGCATCAGTGGCGAAATTTTGTGTCATTTGATGATTCCTGGAAAAAGTATAAAG 720
Qy 241 SerIleLeuAspTyrThrSerPheAsnGlnGluArgIleValAspValAlaGlyProGly 260
Db 721 AGTAICTTGGACTGGACATCTTTTAAACAGGAGAGAAATTTGTTGATGTTGCTGGACCCAGG 780
```


261	Qy	GlyTrpAsnAspProAspMetLeuValIleGlyAsnPheGlyLeuSerTrpAsnGln	280
781	Db	GGTTGGAAATGACCCAGAGATGTTAGTATTGGCAACTTGGCTCAGCTGGAATCAGCAA	840
281	Qy	ValThrGlnMetAlaLeuTrpAlaIleMetAlaAlaProLeuPheMetSerAsnAspLeu	300
841	Db	GTAACCTCAGATGGGCCCTGGGGCTATCATGGCTGCTCCITTTATTCATGCTTAATGACCTC	900
301	Qy	ArgHisIleSerProGlnAlaLysAlaLeuLeuGlnAspLysAspValIleAlaIleAsn	320
901	Db	CGACACATCAGCCCTCAAGCCAAAGCTCTCTCTCAGGATAAGGACGTAATTTGCCATCAAT	960
321	Qy	GlnAspProLeuGlyLysGlnGlyTrpGlnLeuArgGlnGlyAspAsnPheGluValTrp	340
961	Db	CAGAACCCCTTTGGGCACAGCAGGGTACCGACTAGACAGGAGACAACTTTGAAGTGTGG	1020
341	Qy	GluArgProLeuSerGlyLeuAlaTrpAlaValAlaMetIleAsnArgGlnGluIleGly	360
1021	Db	GAAACGACTCTCTCAGGCTTAGCCTGGGCTGTAGCTATGATAAACCGCAGGAGATTGGT	1080
361	Qy	GlyProArgSerThrTrpIleAlaValAlaSerLeuGlyLysGlyValAlaCysAsnPro	380
1081	Db	GGACCTCGCTCTTATACCATCGCAGTTGCTTCCTCGGTAAAGAGAGTGGCCTGTAACTCT	1140
381	Qy	AlaCysPheIleThrGlnLeuLeuProValLysArgLysLeuGlyPheTyrGluTrpThr	400
1141	Db	GCTTGCTTCATCACACAGCTCTCTCCCTGTGAAGAAGGATAGGGTTCTATGAATGGACT	1200
401	Qy	SerArgLeuArgSerHisIleAsnProThrGlyThrValLeuLeuGlnLeu	417
1201	Db	TCAAGGTTAAAGATGCATATAATCCACAGGCACCTGTTTGGCTTCAGCTA	1251

Search completed: January 1, 2006, 03:35:26
Job time : 9545.23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:52 ; Search time 21.0897 Seconds
(without alignments)
148.074 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQRNPELHGCALRFLA.....EWSRLRSHNPTGTVLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 57103 seqs, 7488799 residues

Total number of hits satisfying chosen parameters: 57103

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published Applications AA New.*
- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pap.*
 - 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pap.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pap.*
 - 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pap.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pap.*
 - 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pap.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pap.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	3.7	437	7	US-11-073-626-3
2	82	3.6	437	7	US-11-073-626-1
3	81	3.5	310	7	US-11-074-176-108
4	81	3.5	944	7	US-11-057-058-68
5	80.5	3.5	414	6	US-10-467-657-32
6	80.5	3.5	414	6	US-10-467-657-8372
7	80	3.5	861	7	US-11-038-284-36
8	79.5	3.5	1827	7	US-11-057-058-62
9	79	3.4	322	6	US-10-467-657-8554
10	79	3.4	368	7	US-11-129-143-62
11	78.5	3.4	622	6	US-10-510-386-236
12	78	3.4	760	6	US-10-821-234-1141
13	78	3.4	786	6	US-10-467-962B-103
14	77.5	3.4	522	6	US-10-131-826A-450
15	77.5	3.4	903	7	US-11-057-058-65
16	77	3.4	734	6	US-10-652-893-2
17	76.5	3.3	320	6	US-10-467-657-3254
18	76.5	3.3	717	6	US-10-793-626-3022
19	75.5	3.3	275	7	US-11-055-822-2
20	75.5	3.3	843	6	US-10-645-441-8
21	75	3.3	402	6	US-10-650-326B-21
22	75	3.3	402	7	US-11-051-568-29
23	75	3.3	627	6	US-10-873-528-191
24	74.5	3.2	482	6	US-10-793-626-1314
25	74.5	3.2	573	6	US-10-467-657-684

RESULT 1
US-11-073-626-3
; Sequence 3, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROSAWA, KEIKO
; APPLICANT: HIROKAWA, KOZO
; APPLICANT: KAJIYAMA, NAOKI
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590U50
; CURRENT APPLICATION NUMBER: US/11/073, 626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232, 655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Eupenicillium terrenum
US-11-073-626-3

ALIGNMENTS

26	74.5	3.2	966	7	US-11-057-058-67	Sequence 67, Appl
27	74	3.2	417	6	US-10-467-657-6612	Sequence 6612, Ap
28	73.5	3.2	932	7	US-11-057-058-59	Sequence 59, Appl
29	73	3.2	514	7	US-11-103-037-3	Sequence 3, Appli
30	73	3.2	734	7	US-11-137-465-65	Sequence 65, Appl
31	72	3.1	515	6	US-10-630-203-6	Sequence 6, Appli
32	72	3.1	842	7	US-11-038-284-38	Sequence 38, Appl
33	72	3.1	847	7	US-11-038-284-42	Sequence 42, Appl
34	71.5	3.1	442	6	US-10-485-788A-516	Sequence 516, App
35	70.5	3.1	316	6	US-10-821-234-1526	Sequence 1526, Ap
36	70.5	3.1	834	6	US-10-928-446A-188	Sequence 188, App
37	70.5	3.1	834	6	US-10-928-446A-192	Sequence 192, App
38	70.5	3.1	834	6	US-10-928-446A-194	Sequence 194, App
39	70.5	3.1	834	6	US-10-928-446A-196	Sequence 196, App
40	70.5	3.1	834	6	US-10-928-446A-200	Sequence 198, App
41	70.5	3.1	834	6	US-10-928-446A-202	Sequence 200, App
42	70.5	3.1	834	6	US-10-928-446A-190	Sequence 202, App
43	70.5	3.1	942	6	US-10-928-446A-182	Sequence 182, App
44	70.5	3.1	947	6	US-10-928-446A-184	Sequence 184, App
45	70.5	3.1	950	6	US-10-928-446A-184	Sequence 184, App

Query Match	3.7%	Score 85;	DB 7;	Length 437;
Best Local Similarity	20.4%;	Pred. No. 0.63;		
Matches	46;	Conservative	29;	Mismatches 71;
				Indels 80;
				Gaps 12;
QY	32	LDNLARTPTMGWLHWFMCNLDCEPDSCISEKLFMEAEIMVSEGMKADAGEYLICI	91	
Db	115	LDAGIGLEKTNVWL-----ESEDILAKAPNFTREQV---KGWKG-----LFCT	155	
QY	92	DDCWMAPOBDSRGLQADPQRPFHGIRQLANYVHSGKLKLG-----IYADVGNKT	141	
Db	156	DGGLAAAK-----AINAIGIFLDKGVKFGGAGTFOQPLFAADG-KT	199	
QY	142	CAGFPQSGFYDIDACTFAD-----WGVLLKF-DGC-----YCDLENLADGY	184	
Db	200	CIGLETTD-----TKYFADKVVLAAGAWSPTLVLEDCQVSKAWFAHQLTPEADAY	254	
QY	185	KHMSLALNRTGSRISVYSCWPLYMPPFKPNYTEIRQYCNHWRNFA	230	
Db	255	KNV-----PVVYDGEYGP-----FFEFNEYGVIKVCDEFPGFS	287	

```
RESULT 2
US-11-073-626-1
; Sequence 1, Application US/11073626
; Publication No. US20050244926A1
; GENERAL INFORMATION:
; APPLICANT: KUROKAWA, KEIKO
; APPLICANT: HIROKAWA, NAOKI
; APPLICANT: KAJIYAMA, KOZO
; TITLE OF INVENTION: Novel fructosyl peptide oxidase
; FILE REFERENCE: 227590U80
; CURRENT APPLICATION NUMBER: US/11/073,626
; CURRENT FILING DATE: 2005-03-08
; PRIOR APPLICATION NUMBER: US/10/232,655
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: JP 2001-266665
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: JP 2001-378151
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: JP 2002-228727
; PRIOR FILING DATE: 2002-08-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Coniochaeta sp.
US-11-073-626-1

Query Match      3.6%; Score 82; DB 7; Length 437;
Best Local Similarity 19.5%; Pred. No. 1.2;
Matches 42; Conservative 31; Mismatches 82; Indels 60; Gaps 10;

QY 33 DNGIARTPTMGWLHWFRCMLDCQEEPDSCISEKLFMEAMLMVSEGKWDAGYLYCID 92
Db 116 DAGAGLEKTHAWL-----DNEDEILSKMPLLRDQI---QGKAIWSQ----D 156

QY 93 DCWMAQRDSEGRLOADPQRPFGHIGRLQYLVHSGKGLG-----IYADVGNKTC 142
Db 157 GGLAAAK-----AINAIGQFLKRGVKGFGGAGFKQPLFDDGT-TC 200

QY 143 AGFPGSGFG---YYDIDAQTFADMGVLLKPDGCGDSENLADGKXMSL-----ALNRTG 195
Db 201 IGVETADGTYKYADKKVLAAGASPTLVLEDDQCCSK----AWYAHQILTPEEAAYKG 256

QY 196 RSIVYSCWPLYMPPKPNYTYRQYCNHRNPA 230
Db 257 VPVYNGEFGF----FPEPNEFGVIKVCDEFPGFS 287

RESULT 3
US-11-074-176-108
; Sequence 108, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Kleenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 108
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus

Query Match      3.5%; Score 81; DB 7; Length 310;
Best Local Similarity 20.9%; Pred. No. 1;
Matches 54; Conservative 32; Mismatches 114; Indels 58; Gaps 10;

QY 56 CQEEPDSCISEKLFMEAMLMVSEGKWDAGYLYCIDDCWMAQRDSEGRLOADPQRPFP 115
Db 10 CYNEESI---PLFYNAVE-KVMQTILDDIYEFWFDGSSDSEMLQKNDPEHVV 65

QY 116 GIRQLANYHSGKGLGIYADVGNKTCAGPFGSGFYDIDAQTFADMGVLLKPDGCGYCD 175
Db 66 YVSFSRNFGEKAALYAGLQASIGDIYW-----MDVDLQ-----D 100

QY 176 SLENLADGKXMSLA-----LNRTGSIYVSCWPLYMPPKPNYTYRQY 222
Db 101 PPEFLPEMYNYIKTKGYDCIGTRRVDKGEAKFSFLSDMFYKLVNKNISNTEIVSGARDY 160

QY 223 CNHRNFPAD---IDDSWKSISILDWT9F-----NOERIVDVVAGPGGMNDPMLVI 270
Db 161 RMTROMVDVAVLSMPYENRFSKGFPSWVGFKTKYLDYHNVER---VAGESDWNWTKLKY 217

QY 271 GNFGLS-WNQVVTQMALW 287
Db 218 AMDGIADFSQAPLNIAMV 235

RESULT 4
US-11-057-058-68
; Sequence 68, Application US/11057058
; Publication No. US20050244400A1
; GENERAL INFORMATION:
; APPLICANT: Lebowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-057-058-68

Query Match      3.5%; Score 81; DB 7; Length 944;
Best Local Similarity 26.5%; Pred. No. 4.4;
Matches 36; Conservative 13; Mismatches 49; Indels 38; Gaps 7;

QY 75 LMVSEGKWDAGYLYCIDDCWM-APQRDSEGRLOADPQRPFGH-----IRQLAN----- 122
Db 410 LEVDQGFDD---HNLPCDVIWLDIEHADGKRYFTWDPSPFPQPRMTLERLASKRKLVAI 466

QY 123 -----YVHSGKGLGIYADVGNKTCAG-----FPGSGFYDIDAQTFADMGV 164
Db 467 VDPHIKIVDSGYRVHEELRNGLYV-----KTRDGSDEYGCWPGSAGYPDTNPTMRAWWA 522

QY 165 DLLKFDGCGYCDSENL 180
Db 523 NMFSDYN-YEGSAPNL 537

RESULT 5
US-10-467-657-32
; Sequence 32, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
```

```

QY 137 VGNKTCAGPGSGFYCIDDAQTFADMGVDLLKPKDGCYCDLSLENLADGYKHMSLALNRTGR 195
Db 368 MYNE-IEGWPHA-----EFDG-PCDRLCNLQAEKY-----GK 398

QY 197 SIVYSCWEPLY 207
Db 399 KHYFPCSPLF 409

RESULT 7
US-11-038-284-36
; Sequence 36, Application US/11038284
; Publication No. US20050246793A1
; GENERAL INFORMATION:
; APPLICANT: COOKE, DAVID
; APPLICANT: DEBET, MARTINE
; APPLICANT: GIDLEY, MICHAEL, JOHN
; APPLICANT: JOBLING, STEPHEN, ALAN
; APPLICANT: SAFFORD, RICHARD
; APPLICANT: SIDEBOTTOM, CHRISTOPHER, MICHAEL
; APPLICANT: WESTCOTT, ROGER, JOHN
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO PLANT STARCH COMPOSITION
; FILE REFERENCE: 054163-5003-US
; CURRENT APPLICATION NUMBER: US/11/038,284
; CURRENT FILING DATE: 2005-01-21
; PRIOR APPLICATION NUMBER: US/10/056,454
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: PCT/GB96/01075
; PRIOR FILING DATE: 1996-05-03
; PRIOR APPLICATION NUMBER: GB 9607409.1
; PRIOR FILING DATE: 1996-04-10
; PRIOR APPLICATION NUMBER: GB 9509229.2
; PRIOR FILING DATE: 1995-05-05
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Lathyrus sp.
US-11-038-284-36

Query Match 3.5%; Score 80; DB 7; Length 861;
Best Local Similarity 25.6%; Pred. No. 4.9;
Matches 32; Conservative 17; Mismatches 42; Indels 34; Gaps 6

QY 108 ADPQRF--PHGIRQLANYHVSKGKLL---GIYADVGNKTCAGF-----PGSFG 150
Db 395 APSRFTGFPDLKSLIDRAHELGLLVLDIVHSHSSNNLTDGLNMFDTGDGHYFPGSRG 454

QY 151 YYDI-DAQTF-----ADGVDLLKFDGCYCDLSLENLADGYKHMSLALNRTG 195
Db 455 YHWMDSRLFNYGSWEVLRYLLSNARWLDVEYKFDGFRFDGVTSMW--YTHHGLQVSFTG 512

QY 196 RSIVY 200
Db 513 NYSEY 517

RESULT 8
US-11-057-058-62
; Sequence 62, Application US/11057058
; Publication No. US20050244400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Mega, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3

```

```
; SEQ ID NO 62
; LENGTH: 1827
; TYPE: PRT
; ORGANISM: Rabbit sp.
US-11-057-058-62

Query Match      3.5%; Score 79.5; DB 7; Length 1827;
Best Local Similarity 18.3%; Pred. No. 15;
Matches 42; Conservative 29; Mismatches 61; Indels 97; Gaps 10;

QY 143 AGPFGSYDYDIDQTFADMGV-----DLKFGCYCD-----175
Db 1360 AAPP-----DFFRNSTAEWWTREILDFYNNYKFKDGLWIDMNEPSSFVNGTTTNCVCRNT 1413
QY 176 -----SLENADGYKMSLAIN-----RTGRSIVYSCWPLYMWPQKPNYTEIR- 220
Db 1414 ELNYPYPPELTKRTDGLFRTMCMETEHILSDGSSVLHYDVHNLGYGWSQAKPTYDALQK 1473
QY 221 -----QYCNHW--RNFADIDDSWKSISILDWTSFNQERIVDVAGP- 259
Db 1474 TTGKRGIVISRSYPTAGRWAGWLGNDYARWDMNDSIIGMWFSLFG-----ISYTGAD 1529
QY 260 --CGWNPDMLV-----IGNF-----GLSWNQVQTM 284
Db 1530 ICGFFNDSYHLCTRWTLQLGAFYFARNHNIQTFRRQDPVSWNQTFVEM 1578

RESULT 9
US-10-467-657-8554
; Sequence 8554, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASNIGANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: Seqwin99, version 1.04
; SEQ ID NO 8554
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8554

Query Match      3.4%; Score 79; DB 6; Length 322;
Best Local Similarity 20.7%; Pred. No. 1.7;
Matches 78; Conservative 38; Mismatches 101; Indels 160; Gaps 23;

QY 2 QLRNPEHLGCALALRFALVSWDIPGARALDNGLATPTGWLHWFRCNLDCCQEPD 61
Db 27 QCSVDVHLWLCCEAG-----FAGIARLHPFVKVHVWKWQWRK-----64
QY 62 SCISEKLFMEAMELWSEGWKDG-----YEYLCIDDCWMAQPDSEGRLOA---108
Db 65 -----HLFQAEWTRENGRLKQALLQEAFFDVL-----DSQGLIKSACF 102
QY 109 -----DPORFPHGIRQLA---NYVHSKGLKL-----GIYADVGNKTCAGPGPSF 149
Db 103 AKMAKSPYGLDKHSAREGVAALAYDKYVVPKGRNAVWRNLDFAQV-----F 151
QY 150 GYYDIDAQTFADWGVLLKFDGCGYCDSELENADGYKMSLAINRTGRSIVYSCWPLYM 209
Db 152 GYAMPETQVF---GLSVPE-----ASRLKNLAQPY---YALHATSRD---SKLWPNENW 197
QY 210 P--FQKNPNYTEIROYCN---HWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWND 264
```

```
Db 198 RELLOKLNQ---KQCNNVYLPWGNFAE-----KVRAEQIAD-----230
QY 265 PDMLVIGNFGLSWNQVQTMALWMAAPLF-----MSNDLRHLSPOAKALLODKDV 316
Db 231 -----GLPFTIIVCAKWNL---LQAYLLKHAIVGVDTGLLHL---ANAL---EKPV 274
QY 317 IAI--NQDPLGKQGYQL 331
Db 275 VGIYTDTPDPI-KTGVQV 290

RESULT 10
US-11-129-143-62
; Sequence 62, Application US/11129143
; Publication No. US20050266518A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMBELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISEEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/11/129,143
; CURRENT FILING DATE: 2005-05-13
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Sulfolobus solfataricus
US-11-129-143-62

Query Match      3.4%; Score 79; DB 7; Length 368;
Best Local Similarity 25.6%; Pred. No. 2;
Matches 50; Conservative 32; Mismatches 79; Indels 34; Gaps 11;

QY 176 SLENADGYKMS--LALNRTGRSIVYSCWPLYMWPQKPNYTEIROYCNHWRNPADID 233
Db 175 ALERLRDISKELSVPIIVKESNGI--SMETAKLYSIGIKNFTSGCGGTNWIEMIR 232
QY 234 D----SWK--SIKSILDWTSFNQERIVDVAGPGGWNDPDMLVIGNFGL--SWNQVQTMAL 286
Db 233 DIRRGWKAESAKNFLDWGVPTAAIIIEVR---YSIFDAFLVSGGSGIRSLDAKAAIAL 288
QY 287 WAI---MAAPLFMS-----NDLRHSPOAKA---LLQDKDVAINQDP---LQK-Q 327
Db 289 GADIAGNALPVLKSAIEGKESLEQPFPRKIIFELKATMMLTGSKNVEALKRSSIVILGKLK 348
QY 328 GYQLROGDNFEVWER 342
Db 349 EWABYRGINLSIYEK 363

RESULT 11
US-10-510-386-236
; Sequence 236, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjørke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
; LENGTH: 622
```

```

Db      511 KFQGGQEEERLALE---TALMYGAKKPEL 535
      |||      :|||      :|||
      |||      :|||      :|||

RESULT 13
US-10-467-962B-103
; Sequence 103, Application US/10467962B
; Publication No. US20050246784A1
; GENERAL INFORMATION:
; APPLICANT: Plesch, Gunnar
; APPLICANT: Blau, Astrid
; APPLICANT: Daeschner, Klaus
; APPLICANT: Klein, Mathieu
; TITLE OF INVENTION: Identification of Herbicidally Active Substances
; FILE REFERENCES: 2000 857
; CURRENT APPLICATION NUMBER: US/10/467,962B
; CURRENT FILING DATE: 2003-08-14
; PRIOR APPLICATION NUMBER: PCT/EP02/01466
; PRIOR FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 109
; SOFTWARE: PatentIn Vers. 2.0
; SEQ ID NO 103
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-467-962B-103

Query Match      3.4%; Score 78; DB 6; Length 786;
Best Local Similarity 19.0%; Pred. No. 6.9;
Matches 76; Conservative 39; Mismatches 106; Indels 180; Gaps 20

QY      5 NPETHLGCALALRFLALVSWDIPGARALDNGLAR-----
Db      280 NGYLHGHAKAM-----FVDFGLAKERGGCCYLRYYDTNPAAKEEYINHI 325
      |||      :|||      :|||
      |||      :|||      :|||

QY      39 TPTMGHLHWRFCNLDQCEE-----PDCISE--K 67
      :||      :||      :||
      :||      :||      :||

Db      326 EETVKWGMPEFKELYDLAVELIRGHAYVDHQTADEIKRYREKKOMSPWRDRPIESLK 385
      :||      :||      :||
      :||      :||      :||

QY      68 LFMEMAELMVSEGWKADGVEYLCDDCWMAPOQDSEGRLOADPORFPHGIRQLANTYVHSK 127
      :||      :||      :||
      :||      :||      :||

Db      386 LPDEMGRGIIIEG-----KATLRMKQDMQSDNFMNYDLAY---- 421
      :||      :||      :||
      :||      :||      :||

QY      128 GLKLGTYADVGNKTCAGFPGSFGYYDIDAQTFADMGVDLLKFDGCYCDLSLENLADGYKHM 187
      :||      :||      :||
      :||      :||      :||

Db      422 RIKFAPHPKAGDKWCI-YPS---YD-----YAHCTVDSLENIT---HS 457
      :||      :||      :||
      :||      :||      :||

QY      188 SLALN-RTGRSIVYSEWPL-----YMWPKQPNYTEIRQYCNHWRFAFDIDDSWS 238
      :||      :||      :||
      :||      :||      :||

Db      458 LCTLEFETPRASY--WLLHSLSLYMPYVMEYSRLNVT----- 493
      :||      :||      :||
      :||      :||      :||

QY      239 IKSILDWTSFN---QBRIVDVAGPGWMDPDLVIGNFLGSLWNOQVOTOMALWAIM----- 290
      :||      :||      :||
      :||      :||      :||

Db      494 -NTVMGRKLNLYVTNKYYD-----GWDDPRLITLS--GLR-RRGVSTSTAINAFVRGIGI 544
      :||      :||      :||
      :||      :||      :||

QY      291 ----AAPLFMSNDRHI-----SPQAKALLQDKVDIAIN 320
      :||      :||      :||
      :||      :||      :||

Db      545 TRSDGSMIHVSRLEHHIRELNKTAPRTMVLNPLKVVITN 585
      :||      :||      :||
      :||      :||      :||

RESULT 14
US-10-131-826A-450
; Sequence 450, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey

```

```

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zenin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131.826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 450
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-131-826A-450

Query Match      3.4%; Score 77.5; DB 6; Length 522;
Best Local Similarity 23.3%; Pred. No. 4.5;
Matches 85; Conservative 48; Mismatches 107; Indels 125; Gaps 23;

Qy 102 SEGRLOAPQRFPHGIROL-----ANYVH-----SKGLKGIYADVGN---KTCA 143
Db 145 SYNKLQALAPDLFHLRKLATLHMRANAIOFVPVRFQDCRSILK---FLDIGYNQLKSLA 201
Qy 144 --GPPGSGYDYDIAQTFADWGVDLLKFDGCYCDLSLENLADGKXMSLALNRTGRSIVYS 201
Db 202 RNSFAGLFLKTELHLE-----HNDLVKNFAHPFRLISL-----HSLCRRNKVAIVS 250
Qy 202 C---EWLP-----YWMP--FOKPNVTEIROCNHWRNFAD--IDDSWKSIXSIL 243
Db 251 SLDWVWNLKMDLSGNEIETHEPHVFTVPHLSQLDSDNRLYIEPRILNSKSLTSI- 309
Qy 244 DWTSFNOERIVDVAGPCGWNDPMDLVIGNFGLSMN--QQVTQMALMAIMAAPIFMSN--- 298
Db 310 -----TLAGNL---WDCGRNVCALASW-----LSNFQ 334
Qy 299 -----DLRHISPOAKALLQDKVIT-AINQDPLGKGYQLROGDNFVWVERPLSLGLAWAVM 353
Db 335 RYDGNLQCASPE---YAQGEDVLDVAVYAFHLCEDGAB-----PTSGHLLS-AV 378
Qy 354 INROEIGGRSYTTIYASLKGIV---ACNPACFTITOLLPVKRKLGFEYWTSLRSLHINPT 410
Db 379 TNRSDLGPASSATTIADGEGQHDGTFEPA---TVALP-----GGEHAENAVQIHKVVT 430
Qy 411 GTVLL 415
Db 431 GTNAL 435
```

```

RESULT 15
US-11-057-058-65
; Sequence 65, Application US/11057058
; Publication No. US2005024400A1
; GENERAL INFORMATION:
; APPLICANT: LeBowitz, Jonathan
; APPLICANT: Maga, John
; TITLE OF INVENTION: ACID ALPHA-GLUCOSIDASE AND FRAGMENTS THEREOF
; FILE REFERENCE: SYM-011
; CURRENT APPLICATION NUMBER: US/11/057,058
; CURRENT FILING DATE: 2005-02-10
; PRIOR APPLICATION NUMBER: US 60/543,812
; PRIOR FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 65
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Spinacia oleracea
; US-11-057-058-65
```

```

Query Match      3.4%; Score 77.5; DB 7; Length 903;
Best Local Similarity 17.7%; Pred. No. 9.3;
Matches 84; Conservative 47; Mismatches 136; Indels 207; Gaps 22;

Qy 73 AELMVSEGWKADAGEYLCIDDCWMAQORDSEGRLOADPQRFPP-HGIRQLANYVHSGKLKL 131
Db 343 AKIPLEVMWMTDIDY-----MDAYKDFTLDPVNFPLDKMKKFVNNLHKNGQKY 389
Qy 132 GIYADVGV---NKTGAGFPGSGFYDYDIAQTFADWGVDLLKFDGCYCDLSLENLADGKXMS 188
Db 390 VVILDFGISTNKT-----YETIIRGMKH-D 413
Qy 189 LALNRTGRSIVYSCW--PLYMMPPFOKPN-----YTEIROQCNHWRNFADIDDSWKSIXS 241
Db 414 VFLKRNKGPVYLGSV-WPGVVPYPPDFLKPALSALTFTWTDIQRFL---NLLPVDGLWIDMNE 468
Qy 242 ILDWT-----FNOE 251
Db 469 ISNFISPPIPGSTLDNPPYKINNNGVMLPIINKTIPPTAMHYGDIPEYNNVHNLFGYLEA 528
Qy 252 RIVDVA-----GPG-----GWND-----PDMLVIGNFGLS 276
Db 529 RVTRALILKLEKRPFFVLSRSTFSGSGKYTAHTWTDGNAATWDLVYSIFSLDFGLFGI- 587
Qy 277 WNOQVTOMALWAIMAAPL--FMSNDLRHISPOAKALLQDKVITAINQD--PLGKGYQLR 332
Db 588 -----PMVGADICGFLGNTTEEL---CRRWILQAGAFYFFSRDHSLSLGTYYQELY 633
Qy 333 QGDNFVWVERPLSLGLAWAV-----AMINRQIIGGPRSY-----TIAVAS---L 372
Db 634 RMESVAASARKVLGLRYTLPPYTYLWYEAQLNGIPIARPLFFSFDDDKITGYGISQFLL 693
Qy 373 GKGVACNP-----ACFITQLLP-----VKRKLGFYEWTSLRSLHIN 408
Db 694 GKGVWSPVLKPGVSVSYTAYFPRGNWFDLEDYTRSVTASTGRVYVTLSPDPDHLN 747
```

Search completed: January 1, 2006, 00:29:52
Job time : 22.0897 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:15:26 ; Search time 144.752 Seconds
(without alignments)
2032.482 Million cell updates/sec

Title: US-10-602-220-16
Perfect score: 2293
Sequence: 1 MQLRNPGLHLCALALRFLA.....EWTSLRSHRNPTGVLLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	429	1	P06280 homo sapien
2	2293	100.0	429	1	Q53Y83 HUMAN
3	2292	100.0	429	2	Q53HF3 homo sapien
4	2157	94.1	403	2	O6LER7 homo sapien
5	1806	78.8	419	1	P51569 mus musculus
6	1806	78.8	421	2	Q8BZ66 mus musculus
7	1480.5	64.6	429	2	Q90WL7 FUGRU
8	1267	55.3	377	2	Q4RTE7 TETNG
9	1130	49.3	338	2	Q5XJT6 BRARE
10	1074.5	46.9	431	1	O6Q017 XENLA
11	1056.5	46.1	400	2	Q6GR44 XENLA
12	1051	45.8	405	1	NAGAB_CHICK
13	1049	45.7	412	2	Q7Q1V0 ANOGA
14	1038	45.3	410	2	Q4STX0 TETNG
15	1030	44.9	413	2	Q9VL27 DROME
16	1030	44.9	427	2	Q8MYV3 DROME
17	1007.5	43.9	419	2	Q7Q6H3 ANOGA
18	1006	43.9	417	2	Q9V7N9 DROME
19	990.5	43.2	415	1	NAGAB_MOUSE
20	982	42.8	415	1	NAGAB_RAT
21	967.5	42.2	411	1	NAGAB_BOVIN
22	952.5	41.5	411	1	NAGAB_HUMAN
23	861	37.5	451	2	Q21801_CABEL
24	849	37.0	452	2	Q60L20_CABER
25	759	32.8	399	2	Q5DBS4 SCHJA
26	729	31.8	431	2	Q9FW88 ORYZA
27	716.5	31.2	434	2	Q9LYL2 ARATH
28	716.5	31.2	437	2	Q8VXZ7 ARATH
29	712	31.1	378	1	AGAL_COFAR
30	710	31.0	466	2	Q54EG4 DICDI
31	710	31.0	478	2	Q8VW86_CLOJO

RESULT 1				
ID	AGAL_HUMAN	STANDARD;	PRT;	429 AA.
AC	P06280;			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	01-AUG-1988	(Rel. 08, Last sequence update)		
DT	10-MAY-2005	(Rel. 47, Last annotation update)		
DE	Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-galactoside galactohydrolase) (Alpha-D-galactosidase A) (Agalsidase alfa)			
DE	Name=GLA;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxID=9606;			
OX	[1]			
RN	NUCLEOTIDE SEQUENCE.			
RP	TISSUE=Lymphoblast;			
RC	MEDLINE=89263745; PubMed=2542896;			
RX	Kornreich R., Desnick R.J., Bishop D.F.;			
RA	"Nucleotide sequence of the human alpha-galactosidase A gene.";			
RL	Nucleic Acids Res. 17:3301-3302(1989).			
RL	[2]			
RN	NUCLEOTIDE SEQUENCE.			
RP	TISSUE=Fibroblast;			
RC	MEDLINE=87246603; PubMed=3036505;			
RX	Tuji S., Martin B.M., Kaslow D.C., Migeon B.R., Choudary P.V.,			
RA	Stubblefield B.K., Mayor J.A., Murray G.J., Barranger J.A.,			
RL	Gibbs R.A.;			
RA	"Signal sequence and DNA-mediated expression of human lysosomal alpha-galactosidase A.";			
RL	Eur. J. Biochem. 165:275-280(1987).			
RL	[3]			
RN	NUCLEOTIDE SEQUENCE.			
RP	MEDLINE=95352959; PubMed=7626884;			
RX	Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,			
RA	Gibbs R.A.;			
RL	"Sixty-nine kilobases of contiguous human genomic sequence containing the alpha-galactosidase A and Bruton's tyrosine kinase loci.";			
RL	Mamm. Genome 6:334-338(1995).			
RL	[4]			
RN	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].			
RP	PubMed=15772551; DOI=10.1038/nature03440;			
RX	Ross M.T., Grafham D.V., Coffey A.J., Scherer S., Muzny D.,			
RA	Platzner M., Howell G.R., Burrows C., Bird C.P., Frankish A.,			
RL	Lovell P.L., Howe K.L., Ashurst J.L., Fulton R.S., Sudbrak R., Wen G.,			
RA	Jones M.C., Huttaker M.E., Andrews T.D., Scott C.E., Searle S.,			
RA	Ramsay J., Whitaker A., Deadman R., Carter N.P., Hunt S.E., Chen R.,			
RA	Cree A., Gunaratne P., Havlak P., Hodgson A., Metzker M.L.,			
RA	Richards S., Scott G., Steffen D., Sodergren E., Wheeler D.A.,			
RA	Worley K.C., Ainscough R., Ambrose K.D., Ansari-Lari M.A., Aradhy S.,			
RA	Ashwell R.I., Babbage A.K., Baggeley C.B., Ballabio A., Banerjee R.,			
RA	Barker G.E., Barlow K.F., Barrett I.P., Bates K.N., Beare D.M.,			

32	709	30.9	420	2	Q5DUH8_COFAR	Q5duh8 coffea arab
33	698.5	30.5	425	2	Q7XIV4_ORYZA	Q7xiv4 oryza sativ
34	697.5	30.4	431	2	Q5XTZ3_SALMI	Q5xtz3 salvia milt
35	695.5	30.3	380	2	Q9FUR8_LYCES	Q9fur8 lycopersico
36	695.5	30.3	409	2	Q9SP05_LYCES	Q9sp05 lycopersico
37	694	30.3	378	2	Q5DUH7_COFCA	Q5duh7 coffea cane
38	692	30.2	428	2	Q84VQ7_HELAN	Q84vq7 helianthus
39	687.5	30.0	417	1	AGAL_ORYZA	Q9xtx4 oryza sativ
40	687.5	30.0	422	2	Q39811_SOYBN	Q39811 glycine max
41	684.5	29.9	396	2	Q8RX86_ARATH	Q8rx86 arabidopsis
42	684	29.8	385	2	Q54AX7_DICDI	Q54ax7 dictyosteli
43	682	29.7	425	2	Q41100_PHAVU	Q41100 phaseolus v
44	681.5	29.7	411	1	AGAL_CYATE	P17449 cymopelis c
45	681	29.7	405	2	Q5ZP79_PEA	Q5zpz9 pisum sativ

ALIGNMENTS

RA Beasley H., Beasley O., Beck A., Bethel G., Blechschmidt K., Brady N.,
RA Bray-Allen S., Bridgeman A.M., Brown A.J., Brown M.J., Bonnin D.,
RA Bruford E.A., Buhay C., Burch P., Burford D., Burgess J., Burrill W.,
RA Burton J., Bye J.M., Carder C., Carrel L., Chako J., Chapman J.C.,
RA Chavez D., Chen G., Chen Y., Chen Z., Chinault C.,
RA Ciccodicola A., Clark S.Y., Clarke G., Clee C.M., Clegg S.,
RA Clerc-Blankenburg K., Clifford K., Cobley V., Cole C.G., Conquer J.S.,
RA Corby N., Connor R.E., David R., Davies J., Davis C., Davis J.,
RA Delgado O., Deshazo D., Dhani P., Ding Y., Dinh H., Dodsworth S.,
RA Draper H., Dugan-Rocha S., Dunham A., Dunn M., Durbin K.J., Dutta I.,
RA Eades T., Ellwood M., Emery-Cohen A., Errington H., Evans K.L.,
RA Faulkner L., Francis F., Frankland J., Fraser A.E., Galgoczy P.,
RA Gilbert J., Gill R., Gloeckner G., Gregory S.G., Gribble S.,
RA Gyffiths C., Grocock R., Gu Y., Gwilliam R., Hamilton C., Hart E.A.,
RA Haves A., Heath P.D., Heitmann K., Hennig S., Hernandez J.,
RA Hinzmann B., Ho S., Hoffs M., Howden P.J., Huckle E.J., Hume J.,
RA Hunt P.J., Hunt A.R., Isherwood J., Jacob L., Johnson D., Jones S.,
RA de Jong P.J., Joseph S.S., Keenan S., Kelly S., Kershaw J.K., Khan Z.,
RA Xiochis P., Klages S., Knights A.J., Kosiura A., Kovar-Smith C.,
RA Laird G.K., Langford C., Lawlor S., Leversha M., Lewis L., Liu W.,
RA Lloyd C., Lloyd D.M., Loubeed H., Loveland J.E., Lovell J.D.,
RA Lozado R., Lu J., Lyne R., Ma J., Maheshwari M., Matthews L.H.,
RA McDowall J., McLaren S., McMurray A., Meidl P., Meitinger T.,
RA Milne S., Miner G., Mistry S.L., Morgan M., Morris S., Mueller I.,
RA Mullikin J.C., Nguyen N., Nordsiek G., Nyakatura G., O'dell C.N.,
RA Okwou G., Palmer S., Pandian R., Parker D., Parrish J.,
RA Pasternak S., Patel D., Pearce A.V., Pearson D.M., Pelan S.E.,
RA Perez L., Porter K.M., Ramsey Y., Reichwald K., Rhodes S.,
RA Ridler K.A., Schlesinger D., Schueler M.G., Sehra H.K.,
RA Shaw-Smith C., Shen H., Sheridan E.M., Showken R., Skuce C.D.,
RA Smith M.L., Sothran E.C., Steingrubber H.E., Steward C.A., Storey R.,
RA Swann R.M., Swardbeck D., Tabor P.E., Taudien S., Taylor T.,
RA Teague B., Thomas K., Thorpe A., Timms K., Tracey A., Trevanion S.,
RA Tromans A.C., d'Urso M., Verdusco D., Villaseana D., Waldron L.,
RA Wall M., Wang Q., Warren J., Warry G.L., Wei X., West A.,
RA Whitehead S.L., Whiteley M.N., Wilkinson J.E., Willey D.L.,
RA Williams G., Williams L., Williamson A., Williamson H., Wilming L.,
RA Woodmansey R.L., Wray P.W., Yen J., Zhang J., Zhou J., Zoghbi H.,
RA Zorilla S., Buck D., Reinhardt R., Pouster A., Rosenthal A.,
RA Leirach H., Meindl A., Minx P.J., Hillier L.W., Willard H.F.,
RA Wilson R.K., Waterston R.H., Rice C.M., Vaudin M., Coulson A.,
RA Nelson D.L., Weinstein G., Suleston J.E., Durbin R., Hubbard T.,
RA Gibbs R.A., Beck S., Rogers J., Bentley D.R.;
RT "The DNA sequence of the human X chromosome.";
RL Nature 434:325-337(2005).
[5]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC TISSUE=Uterus;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.W., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[6]
RP NUCLEOTIDE SEQUENCE OF 31-429, AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.F., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Deanick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863 (1986).
[7]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=98112869; PubMed=2892762; DOI=10.1016/0378-1119(87)90374-X;
RA Quinn M., Hantzopoulos P., Fidanza V., Calhoun D.H.;
RT "A genomic clone containing the promoter for the gene encoding the
human lysosomal enzyme, alpha-galactosidase A.";
RL Gene 58:177-188 (1987).
[8]
RP NUCLEOTIDE SEQUENCE OF 1-64.
RX MEDLINE=98234528; PubMed=2836863;
RA Bishop D.F., Kornreich R., Deanick R.J.;
RT "Structural organization of the human alpha-galactosidase A gene:
further evidence for the absence of a 3' untranslated region.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:3903-3907 (1988).
[9]
RP RNA EDITING OF POSITION 396.
RX MEDLINE=95380278; PubMed=7503918;
RA Novo F.J., Kruszewski A., McDermot K.D., Goldspink G., Gorecki D.C.;
RT "Editing of human alpha-galactosidase RNA resulting in a pyrimidine to
purine conversion.";
RL Nucleic Acids Res. 23:2636-2640 (1995).
[10]
RP X-RAY CRYSTALLOGRAPHY (3.45 ANGSTROMS) OF 32-422 IN COMPLEX WITH
PRODUCT, HOMODIMERIZATION, AND N-GLYCOSYLATION.
RX PubMed=15003450; DOI=10.1016/j.jmb.2004.01.035;
RA Garman S.C., Garboczi D.N.;
RT "The molecular defect leading to Fabry disease: structure of human
alpha-galactosidase.";
RL J. Mol. Biol. 337:319-335 (2004).
[11]
RP REVIEW ON FD VARIANTS.
RX MEDLINE=94258158; PubMed=7911050;
RA Eng C.M., Deanick R.J.;
RT "Molecular basis of Fabry disease: mutations and polymorphisms in the
human alpha-galactosidase A gene.";
RL Hum. Mutat. 3:103-111 (1994).
[12]
RP VARIANT FD SER-40.
RX MEDLINE=90092580; PubMed=2152885; DOI=10.1016/0014-5793(90)80046-L;
RA Koide T., Ishiura M., Iwai K., Inoue M., Kaneda Y., Okada Y.,
RA Uchida T.;
RT "A case of Fabry's disease in a patient with no alpha-galactosidase A
activity caused by a single amino acid substitution of Pro-40 by
Ser.";
RL FEBS Lett. 259:353-356 (1990).
[13]
RP VARIANT FD VAL-296.
RX MEDLINE=91101674; PubMed=1846223;
RA von Scheidt W., Eng C.M., Fitzmaurice T.F., Erdmann E., Hubner G.,
RA Olsen E.G.J., Christomanou H., Kandolf R., Bishop D.F., Deanick R.J.;
RT "An atypical variant of Fabry's disease with manifestations confined
to the myocardium.";
RL N. Engl. J. Med. 324:395-399 (1991).
[14]
RP VARIANT FD GLN-301.
RX MEDLINE=91022721; PubMed=2171331;
RA Sakuraba H., Oshima A., Fukuhara Y., Shimamoto M., Nagao Y.,
RA Bishop D.F., Deanick R.J., Suzuki Y.;
RT "Identification of point mutations in the alpha-galactosidase A gene
in classical and atypical hemizygotes with Fabry disease.";
RL Am. J. Hum. Genet. 47:784-789 (1990).
[15]
RP VARIANT FD TRP-356.
RX MEDLINE=89198098; PubMed=2539398;
RA Bernstein H.S., Bishop D.F., Astrin K.H., Kornreich R., Eng C.M.,
RA Sakuraba H., Deanick R.J.;
RT "Fabry disease: six gene rearrangements and an exonic point mutation
in the alpha-galactosidase gene.";

```

RL J. Clin. Invest. 83:1390-1399(1989).
RN [16]

Query Match      100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 7.5e-186;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSEGRLOADPORPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSEGRLOADPORPHGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDDAQTADFADGWVLLKFDGICYCDSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDDAQTADFADGWVLLKFDGICYCDSLENL 180

Qy 181 ADGKXMSLALNRGSRIVYSCWPLVYMWFPQKPNYTEIROYCNHWRNFADIDDSWKSIX 240
Db 181 ADGKXMSLALNRGSRIVYSCWPLVYMWFPQKPNYTEIROYCNHWRNFADIDDSWKSIX 240

Qy 241 SILDWTSENGRERIVDVAGPGGNDPDLVIGNFGLSWQVQVOTQALWAIMAAPLFMSNDL 300
Db 241 SILDWTSENGRERIVDVAGPGGNDPDLVIGNFGLSWQVQVOTQALWAIMAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVITAINQDPLGKQGYQLRQGDNFVWERPISGLAWAVAMINRQIG 360
Db 301 RHISPOKALLQDKDVITAINQDPLGKQGYQLRQGDNFVWERPISGLAWAVAMINRQIG 360

Qy 361 GPRSYTTIYAVASLKGVCNACPFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 GPRSYTTIYAVASLKGVCNACPFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 2
Q53Y83 HUMAN
ID Q53Y83 HUMAN PRELIMINARY; PRT; 429 AA.
AC Q53Y83
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Galactosidase, alpha.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Maruyama K., Sugano S.;
SU Suzuki Y., Yoshitomo K., Maruyama K., Sugano A., Sugano S.;
RT "Construction and characterization of a full length-enriched and a 5'-
end-enriched cDNA library.";
RL Gene 200:149-156(1997).
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Cerebellum;
RA Suzuki Y., Sugano S., Totoki Y., Toyoda A., Takeda T., Sakaki Y.,
RA Tanaka A., Yokoyama S.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK22627; BAD96347.1; -; mRNA.
SQ SEQUENCE 429 AA; 48752 MW; 613F8BF21B107D7B CRC64;

Query Match      100.0%; Score 2293; DB 2; Length 429;
Best Local Similarity 100.0%; Pred. No. 7.5e-186;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDCCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSEGRLOADPORPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGHKADAGEYLCIDDCWMAPOQDSEGRLOADPORPHGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDDAQTADFADGWVLLKFDGICYCDSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDDAQTADFADGWVLLKFDGICYCDSLENL 180

```

```
QY 181 ADGKXKSLALNRTGRSIVYCEWPLVYMPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
DB 181 ADGKXKSLALNRTGRSIVYCEWPLVYMPFQKPNYTEIROQCNHWRNFADIDDSWKSIIK 240
QY 241 SILDWTSFNQRIYDVVAGPGGWNDDPMLVIGNFGLSNQVQVQTMALWAIWAAPLPMNSNDL 300
DB 241 SILDWTSFNQRIYDVVAGPGGWNDDPMLVIGNFGLSNQVQVQTMALWAIWAAPLPMNSNDL 300
QY 301 RHISFOKALLQDKVITAINQDPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINRQIEG 360
DB 301 RHISFOKALLQDKVITAINQDPLGKQGYQLRQGDNFVWERPLSLGLAWAVAMINRQIEG 360
QY 361 GPRSYTTAVASLGKGVACNPACFTTQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417
DB 361 GPRSYTTAVASLGKGVACNPACFTTQLLPVKRKLGFYEWTSLRSHINPTGTVLQL 417

RESULT 4
Q6LER7_HUMAN
ID Q6LER7_HUMAN PRELIMINARY; PRT; 403 AA.
AC Q6LER7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Alpha-galactosidase A (EC 3.2.1.22) (Fragment).
GN Name=alpha-GalA;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=86259694; PubMed=3014515;
RA Bishop D.P., Calhoun D.H., Bernstein H.S., Hantzopoulos P., Quinn M.,
RA Deanick R.J.;
RT "Human alpha-galactosidase A: nucleotide sequence of a cDNA clone
RT encoding the mature enzyme.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4859-4863 (1986).
DR EMBL; D00039; BAA34059.1; -; mRNA.
DR SMR; Q6LER7; 6-396.
DR GO; GO:0004557; F:alpha-galactosidase activity; IEA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR000111; Glyco_hydro_27.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD02572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
KW Glycosidase; Hydrolase.
FT CHAIN 1..403 alpha-galactosidase A subunit peptide.
FT NON TER 1
SQ SEQUENCE 403 AA; 45804 MW; 2FE193205BEB8D1A CRC64;

Query Match 94.1%; Score 2157; DB 2; Length 403;
Best Local Similarity 100.0%; Pred. No. 2.4e-174;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 PGARALDGLARTPTMGWLHWRPFCNLDCCQEEPDCISEKLFMEVABLWVSEGWKDAY 86
DB 1 PGARALDGLARTPTMGWLHWRPFCNLDCCQEEPDCISEKLFMEVABLWVSEGWKDAY 60
QY 87 EYLCDIDCWMAQRDSEGRQLQADQRFPHGIRQLANYVHSGKLGIVADVGNKTCAGFP 146
DB 61 EYLCDIDCWMAQRDSEGRQLQADQRFPHGIRQLANYVHSGKLGIVADVGNKTCAGFP 120
QY 147 GSGFYDIDAQTFADWGVDLKFKPGCCDLSLENLADGKXKSLALNRTGRSIVYCEWPL 206
DB 121 GSGFYDIDAQTFADWGVDLKFKPGCCDLSLENLADGKXKSLALNRTGRSIVYCEWPL 180
QY 207 YMPFQKPNYTEIROQCNHWRNFADIDDSWKSIIKSIDWTSFNQRIYDVVAGPGGWNDDP 266
```

RESULT 5

```
AGAL MOUSE
ID _AGAL_MOUSE STANDARD; PRT; 419 AA.
AC P51569;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Alpha-galactosidase A precursor (EC 3.2.1.22) (Melibiase) (Alpha-D-
DE galactoside galactohydrolase) (Alpha-D-galactosidase A).
GN Name=Gla; Synonyms=Ags;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=96125203; PubMed=8543175; DOI=10.1016/0378-1119(95)00592-7;
RA Oshlma T., Murray G.J., Nagle J.W., Quirk J.M., Kraus M.H.,
RA Barton N.W., Brady R.O., Kulkarni A.B.;
RT "Structural organization and expression of the mouse gene encoding
RT alpha-galactosidase A.";
RL Gene 166:277-280 (1995).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C129;
RX MEDLINE=95352959; PubMed=7626884;
RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D.M., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci.";
RL Mamm. Genome 6:334-338 (1995).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96316016; PubMed=8733892; DOI=10.1006/bmme.1996.0020;
RA Gotlib R.W., Bishop D.F., Wang A.M., Zeidner K.M., Ioannou Y.I.,
RA Adler D.A., Diatche C.M., Desnick R.J.;
RT "The entire genomic sequence and cDNA expression of mouse alpha-
RT galactosidase A.";
RL Biochem. Mol. Med. 57:139-148 (1996).
RN [4]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC STRAIN=FVB/N; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Straube R.D., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heih F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
```


DR ZFIN; ZDB-GENE-041010-207; zgc:101584.
DR GO: 0004553; P: hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: 0005975; P: carbohydrate metabolism; IEA.
DR InterPro: IPR002241; Glyco_hydro_27.
DR InterPro: IPR000111; Glyco_hydro_GHD.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco hydro GHD; 1.
SQ SEQUENCE 338 AA; 37846 MW; A3FC5C461B7C9649 CRC64;

Query Match 49.3%; Score 1130; DB 2; Length 338;
Best Local Similarity 60.3%; Pred. No. 2.8e-87;
Matches 210; Conservative 47; Mismatches 75; Indels 16; Gaps 2;

QY 68 LFEMAEALMVSEGWKADAGEYLCLDDCWAPQRDSEGRQLQADPQRPFGIRQLANYVHSK 127
DB 2 LQMSLCALMIA-LGRNG-----THRGACRQTPKGFPPVASKKLLADYVHSK 45

QY 128 GLKGLGIYADVGNKTCAGFPFGSYDDIDAQTFADGWGVDLLKFDGCYCDLSLENLADGKYM 187
DB 46 GLKGLGIYADVGNKTCAGFPFGSLGYDDIDAQTFADGWGVDLLKFDGCFMPDWHQLGEGYIM 105

QY 188 SLALNRTGRSIVSCEWPLYMPPKPNYTYEIRQYCNHWRNFADIDDSWKSISILDWTS 247
DB 106 SSALNRTGRSIVSCEWPLYEWHQPDYEIRKTCNHRNRYGVDYQWTSVKSLDWT 165

QY 248 FNQERIVDVAGPGGNDPDLVGNFGLSNQVQTMALWALMAAPLFMSNDLRLHISPOA 307
DB 166 EKQKIVVPVAGPGGNDPDLVGNFGLSNQVQTMALWALMAAPLFMSNDLRLDIPC 225

QY 308 KALLQDKVDIAINQDPLGKQYQLRQGNFVWERPLSGLAWAMINRQISGPRSYTI 367
DB 226 KELLQNKQIAINQDPLGKQYRLKADSFELWERPLSGNLRAVAVNRQISGPRFTI 285

QY 368 AVASLGKGVACNPAFCITQLLPVVRKGLGFYEWTSRLASHNPTGVLL 415
DB 286 SVAIMPSWKLNCPRKCNVQILPTKYKENGQVLLSEVVVVQVNPQTGTTLL 333

RESULT 10
Q6GQ17_XENLA PRELIMINARY; PRT; 431 AA.
AC Q6GQ17;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE LOC443592 protein (fragment).
GN Name=LOC443592;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Embryo;
RC MEDLINE=22348257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzanski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.,
RA "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Embryo;
RA Klein S., Gerhard D.S.; to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072931; AAH72931.1; -; mRNA.
DR GO: 0004553; P: hydrolase activity, hydrolyzing O-glycosyl . . . ; IEA.
DR GO: 0005975; P: carbohydrate metabolism; IEA.
DR InterPro: IPR002241; Glyco_hydro_27.
DR InterPro: IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; 1.
DR PRINTS; PR00740; GLHYDRLASE27.
DR ProDom; PD002572; Glyco hydro GHD; 1.
DR PROSITE; PS00512; ALPHA_GALACTOSIDASE; 1.
FT NON TER 1
SQ SEQUENCE 431 AA; 48787 MW; 4C578DB433983471 CRC64;

Query Match 46.9%; Score 1074.5; DB 2; Length 431;
Best Local Similarity 49.3%; Pred. No. 1.9e-82;
Matches 209; Conservative 62; Mismatches 126; Indels 27; Gaps 6;

QY 4 RNFELHLCALALRFLALVSNDFPGARALDGLARTPTMGLWHERFMCNLDCCOEPDSC 63
DB 18 RGYRMHLTSLCLLTLMLGLCW-----CLDNLGLVTPPMGMMTWQTRCIDCKSDPDNC 71

QY 64 ISEKLFMEAEALMVSEGWKADAGEYLCLDDCWAPQRDSEGRQLQADPQRPFGIRQLANY 123
DB 72 ISENLIKIMADEMADSCWRDLGYVYVICDDCWQKQKQDSNGRLQDPDPFPPSGMKALADY 131

QY 124 VHSKGLKGLIYADVGNKTCAGFPFG-SFGYYDDIDAQTFADGWGVDLLKFDGCYCDLSLENLAD 182
DB 132 VHAKGLKGLIYSDMGNYTCGYPGTTLDTIKTDAETPASWEVDMLKFDGCYSNSTK-AL 190

QY 183 GYKXMSLALNRTGRSIVSCEWPLYMPPF-QKNYTYEIRQYCNHWRNFADIDDSWKSIS 241
DB 191 GYPRMSALNATGRPILYSCSWPAYEGGLPPKVNYYTQLGICNMWRNYDDIDDSWSDVLD 250

QY 242 ILDWTSTNQERIVDVAGPGGNDPDLVGNFGLSNQVQTMALWALMAAPLFMSNDLRL 301
DB 251 IIEWAKNQDVLQPAAGPGHNDPDLITGDFGLSYEQSKSQALAIWAILAAPLMSNDLRL 310

QY 302 HISPOAKALLQDKVDIAINQDPLGKQYQLRQGNFVWERPLSGLAWAMINRQISG 361
DB 311 AISQDAKDLLQNRLLIYINQDSLKQKQSLISRGLENLEWVKRELINGQYALAVLNKGTDL 370

QY 362 PRSVTIASVSLGKGVACNPAFCITQLLP-----VKRKLGFYEWTSRLASHNPTGV 413
DB 371 PRPYSTNLGLLN-----VTQDGYKLYNVFEKYLGMFTSTPTPIIRVNPQVI 420

QY 414 LLQL 417
DB 421 FLFL 424

RESULT 11
Q6GR44_XENLA PRELIMINARY; PRT; 400 AA.
AC Q6GR44;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)


```

DE GN MGC81044 protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP TISSUE=Ovary.
RC MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RX MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Ovary;
RA Klein S., Gerhard D.S.;
RL Submitted (May-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC071089; AAH71089.1; -. mRNA.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.
DR Pfam; PF02065; Melibiase; I.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA-GALACTOSIDASE; 1.
SQ SEQUENCE 400 AA; 44854 MW; CBC4770F613FF3BA CRC64;

Query Match 46.1%; Score 1056.5; DB 2; Length 400;
Best Local Similarity 51.0%; Pred. No. 5.8e-81;
Matches 201; Conservative 58; Mismatches 114; Indels 21; Gaps 5;

Qy 32 LDNGARTPTMGWHLWRFMNCNLCQBPDSICSEKLFMEAEMLVMSGKWDAGYEYLCI 91
Db 7 LDNGLVTPPMGWTWQRYRCNICKGDPDNCISENLKSMADKNADSGWRDVGTYTISI 66

Qy 92 DDCWMAQRDSEGLQADPQPFPHGIRQLANYVHSKGLGIYADVGNKTCAGPPG-SFG 150
Db 67 DDCWSQQRDSNGRLQDPLERFPFGMKALADYVHAKGLGIYDGMGTTCGGYPGGTLD 126

Qy 151 YYDIDAOTFADGWGVDLLKFDCCYCDLSLENLADGKHKMSLALNRTGRSVSCSEPLYMWP 210
Db 127 TIKIDAETFAFWEDMLKFDCCYCSNSTEK-ALGPKMSSEALNGTRPILYSCSWPAYEGG 185

Qy 211 F-QKPNYTEIRQYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGGWNDDPMLV 269

```

```

Db 186 LPPKVNVTQSGICNMWRNTYGDIDSDVSDIIEWYAKNQDVLQPAAGPGRWNPDKMLI 245
Qy 270 IGNFGLSWNQVQTOMALWMAAPLMSNDLRHISPOKALLQDKVIAINQDPLGKQGY 329
Db 246 TGFGLSYEQSKQLAIWAILAAPLINSNDLRITISQADKOLLQNRLLIYNQDALGKQGS 305
Qy 330 QLRQDNFVWERPLSLGAWAVAMINRQETGGPRSYTIVASLKGKGVACNPACFITQLLP 389
Db 306 LISQVGLVYKRELINGQYAVAVLNKGTDLGPRYTTTSLGLN-----ITQCTD 355
Qy 390 -----VRKGLGYEWTSLRSHINPTGTVLL 415
Db 356 GYKMYNVFEKEYIGMFKSGTPIDMRVNPVTGVIFL 389

RESULT 12
NAGAB_CHICK STANDARD; PRT; 405 AA.
ID NAGAB_CHICK
AC Q90744;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Alpha-N-acetylgalactosaminidase (EC 3.2.1.49) (Alpha-galactosidase B).
GN Name=NAGA;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE [MRNA], AND PARTIAL PROTEIN SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=94060104; PubMed=8241271; DOI=10.1016/0167-4781(93)90158-A;
RA Davis M.O., Hata J., Smith D., Walker J.C.;
RT "Cloning and sequence of a chicken alpha-N-acetylgalactosaminidase
gene.";
RL Blochim. Biophys. Acta 1216:296-298 (1993).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX PubMed=12005440; DOI=10.1016/S0969-2126(02)00726-8;
RA Garman S.C., Hannick L., Zhu A., Garbozsi D.N.;
RT "The 1.9 A structure of alpha-N-acetylgalactosaminidase: molecular
basis of glycosidase deficiency diseases.";
RL Structure 10:425-434 (2002).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of terminal non-reducing N-acetyl-
D-galactosamine residues in N-acetyl-alpha-D-galactosaminides.
CC -|- SUBCELLULAR LOCATION: Lysosomal (By similarity)
CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 27 family.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC EMBL; L18754; AAA16614.1; -. mRNA.
DR PIR; S45522; S45522.
DR PDB; 1KTB; X-ray; A=1-405.
DR PDB; 1KTC; X-ray; A=1-405.
DR Ensembl; ENSGALG0000011900; Gallus gallus.
DR InterPro; IPR002241; Glyco_hydro_27.
DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; I.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
DR PROSITE; PS00512; ALPHA-GALACTOSIDASE; 1.
KW 3D-structure; Direct protein sequencing; Glycoprotein; Glycosidase;
Hydrolase; Lysosome.
FT CHAIN 1 405 Alpha-N-acetylgalactosaminidase.
FT ACT_SITE 140 140 Nucleophile (By similarity).
FT ACT_SITE 201 201 Proton donor (By similarity).
FT CARBOHYD 161 161 N-linked (GlcNAc...).
```

```
FT CARBOHYD 185 185 N-linked (GlcNAc. . .).
FT CARBOHYD 369 369 N-linked (GlcNAc. . .).
FT DISULFID 21 63
FT DISULFID 25 32
FT DISULFID 111 142
FT DISULFID 171 193
SQ SEQUENCE 405 AA; 45615 MW; E1EC0061739C305C CRC64;

Query Match 45.8%; Score 1051; DB 1; Length 405;
Best Local Similarity 53.0%; Pred. No. 1.7e-80;
Matches 206; Conservative 58; Mismatches 117; Indels 8; Gaps 6;

Qy 32 LDNGLARTPTMGWLHWRFCMCLDCEEPDSCISEKLFMEAEIWMVSEGWDAGYEYLCI 91
Db 1 LENGARTPPMGWLAWERFCNVCRDPDPCISSEMLFMEADRIADGRELGYKINI 60

Qy 92 DDCWMAQRDSEGRLOADPQRPFHGIRQLANYVHSGKLGKIYADVGNKTCAGPPG-SFG 150
Db 61 DDCWAAQRDAEGRVDPDPERFPRGIRKALADYVHARGLKLDIYDGLGRLTCGGYPGTTLD 120

Qy 151 YYDIDAQTFADWGVLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLYMWP 210
Db 121 RVEQDAQTFADWGVLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLYMWP 210

Qy 211 F-QRPNTYETROYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNPDMVLV 269
Db 180 LPPKNTYTLGEICNLWRNYDDIQDSWDSVLSIVDWFYTDQDVLPQFAGFGHWNPDMLI 239

Qy 270 IGNFGLSNQOQVTOMALWMAAPLPMNSDLRHISPOAKALLQDKOVIAINQDPLGKGY 329
Db 240 IGNFGLSYEQSRQSMALWTMAAPLPMNSDLRHISPOAKALLQDKOVIAINQDPLGKGY 329

Qy 330 Q-LRQGNFVFWERPLSGLAWAMINROEIGGPRSYTIAVASLGKGVACNPACFITOLL 388
Db 300 RIIEGSHIEVFLRPLSQASALVFFSRR-TDMPFRYTTSLAKLGFPWG---AAVEYQDV 355

Qy 389 PVKKEKLGFEWTSRLRSHINPTGTLLQ 417
Db 356 YSGKIISGLKGDNTVIINPFGVVMWYL 384

RESULT 13
Q01V0 ANOGA PRELIMINARY; PRT; 412 AA.
AC Q01V0;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ENSANGP0000020847 (Fragment).
GN ORFNames=ENSANGG0000018358;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Prerygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG "Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008980; EAL14548.2; -; Genomic DNA.
DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR002241; Glyco_hydro_27.

DR InterPro; IPR000111; Glyco_hydro_GHD.
DR Pfam; PF02065; Melibiase; I.
DR PRINTS; PR00740; GLHYDRASE27.
DR PRODOM; PD002572; Glyco_hydro_GHD; 1.
FT NON TER 1
FT NON TER 412
SQ SEQUENCE 412 AA; 46517 MW; 4C91C8FB9EA6C88F CRC64;

Query Match 45.7%; Score 1049; DB 2; Length 412;
Best Local Similarity 49.2%; Pred. No. 2.6e-80;
Matches 193; Conservative 71; Mismatches 118; Indels 10; Gaps 4;

Qy 31 ALDNGLARTPTMGWLHWRFCMCLDCEEPDSCISEKLFMEAEIWMVSEGWDAGYEYLC 90
Db 17 SLENGLARTPPMGWLAWERFCNVCRDPDPCISSEMLFMEADRIADGRELGYKINI 76

Qy 91 IDCWMAQRDSEGRLOADPQRPFHGIRQLANYVHSGKLGKIYADVGNKTCAGPPGSG 150
Db 77 VDCWLEKSRGPRGELVADRRRPFSGMKALANTVHAKLGFYEDYGNVTCAGYPGILG 136

Qy 151 YYDIDAQTFADWGVLLKPDGCGYCDLSLENLADGKHSALNRTGRSIVVSCWPLY-MW 209
Db 137 FSANDAAQFASWVDVYVYKDGCSLPID-MDHGYPEPGRNLNATGRPMIYSCSNPVYQIY 195

Qy 210 PFQKPNYETROYCNHWRNFADIDDSWKSISILDWTSFNQERIVDVAGPGWNPDMVLV 269
Db 196 AGNPNYSSIIQHNLWRNYDDIQDSWASLESIIIDYGNQDALIPNAGFGHWNPDMLI 255

Qy 270 IGNFGLSNQOQVTOMALWMAAPLPMNSDLRHISPOAKALLQDKOVIAINQDPLGKGY 329
Db 256 IGNFGLSYEQSKTOMALWMAAPLPMNSDLRHISPOAKALLQDKOVIAINQDPLGKGY 329

Qy 330 QLRQGNFVFWERPLSGLAWAMINROEIGGPRSYTIAVASLGKGVACNPACFI 384
Db 316 RIYKHGIEIWSRPITPIYQTYTYSYAFNRRRTDGTSPDVAVTLRELG---LISPTGYR 372

Qy 385 TOLLPVKKEKLGFEWTSRLRSHINPTGTLLQ 416
Db 373 VEDLYEVDYGVLSQTKIKVKNPFGVILR 404

RESULT 14
Q4STX0 TETNG PRELIMINARY; PRT; 410 AA.
AC Q4STX0;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAF14084, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00012743001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
[1]
RN NUCLEOTIDE SEQUENCE.
RP Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Blemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouz J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McSwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
```

```

RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAAB01014084; CAP9512.1; -; Genomic_DNA.
FT NON TER 410
SQ SEQUENCE 410 AA; 46172 MW; F5C661DA72DCEB88 CRC64;

Query Match 45.3%; Score 1038; DB 2; Length 410;
Best Local Similarity 48.9%; Pred. No. 2.2e-79;
Matches 207; Conservative 58; Mismatches 116; Indels 42; Gaps 8;

Qy 8 LHGLCALALFLALVSWDIPCARALDNGLARTPTMGWLHWRFCNLDCEEPDSCISEK 67
Db 1 MHLALLLASVLSL-----GTLALDNGLWRTPTMGWLHWRFCNLDCEEPDSCISEN 54
Qy 68 LFMEAMELVSEGKADAGBYLCIDDCWMAQPDSEGRQLQADPQRPFGIRQLANYVHSK 127
Db 55 LFTDMADRLSQDGKWDGLGVYVYVNDICWSSKERDKGRLQDPDKRPPGGIPKLARYMHR 114
Qy 128 GLKGIYADVGNKTCAGFPQS-FGYDIDAQTPADWGVLLKFDGCVCDSENLADGYKH 186
Db 115 GLKGIYDGMGTLTCGGYPTGLDKIDIDAKTFAEMEVDMLKFDGCVSNEVEQ-QQGYPL 173
Qy 187 MSALNATGRSIVYSCWPLVMPWF-QKPNYETIRQYCNHWRNPADIDDSWKSITLDW 245
Db 174 MSXALNATGRPIAYSCSWPAYSGGLPPKVNVTOLGEICNLWRNYDDIQDSWDSVNLIDW 233
Qy 246 TSFNQERIVDVAGPGWNDPDLVIGFGLSWNQVQTQMALWAIMAPLFWNSDLRHIS 305
Db 234 FFDNQDVIAAPAGGRWNDPDLVIGFGLSLQSRQTQMALWAIMAPLFWNSDLRTISA 293
Qy 306 QAKALLQKDVIAINQDPLGKQVQL-RQGNFVFWERPLUGSLAWAVAMINROEIGPRS 364
Db 294 EARSILQKIAIGINQDPLGQRRLLVKEKSGIEVFWRLSDNTSALVFFSR----- 346
Qy 365 YTTAVASLGKGVACNPACTTOLLVPRKLGFE-----WTSRLRS-----HINPTG 411
Db 347 -----TDPYRYKTSLSKUSYPTGYKIVDVFTQKSSLSKDSADFVVSVPNTG 394

Qy 412 TVL 414
Db 395 VVM 397

RESULT 15
Q9VL27 DROME
ID Q9VL27 DROME PRELIMINARY; PRT; 413 AA.
AC Q9VL27;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CG5731-PA.
GN Names=CG5731; ORFNames=CG5731;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```

```

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry K.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J.W., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084.1-RESEARCH0084.20(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RG Berkeley Drosophila Genome Project;
RA Celniker S., Carlson J., Wan K., Pfeiffer B., Frise E., George R.,
RA Hoskins R., Stapleton M., Pacleb J., Park S., Svirkas R., Smith E.,
RA Yu C., Rubin G.;
RT "Drosophila melanogaster release 4 sequence."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 21:05:01 ; Search time 152.421 Seconds
(without alignments)
1202.074 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNPELHGICALALRFLA.....EWTSLRSHINPTGTVLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	417	5	Aae28212 Human rga
2	2293	100.0	417	7	Add84755 Human alp
3	2293	100.0	417	8	Adj88281 Human wt
4	2293	100.0	417	8	Adm48685 Human wil
5	2293	100.0	417	8	Adu65920 Human alp
6	2293	100.0	417	9	Aae27449 Human alp
7	2293	100.0	421	5	Aae28210 Human rga
8	2293	100.0	421	7	Add84751 Human alp
9	2293	100.0	421	8	Adj88277 Human wt
10	2293	100.0	421	8	Adm48681 Human wil
11	2293	100.0	421	8	Adu65916 Human alp
12	2293	100.0	421	9	Aae27445 Human alp
13	2293	100.0	423	5	Aae28213 Human rga
14	2293	100.0	423	7	Add84757 Human alp
15	2293	100.0	423	8	Adj88283 Human wt
16	2293	100.0	423	8	Adm48687 Human wil
17	2293	100.0	423	8	Adu65922 Human alp
18	2293	100.0	423	9	Aae27451 Human alp
19	2293	100.0	425	9	Aae27472 Human alp
20	2293	100.0	427	5	Aae28211 Human rga
21	2293	100.0	427	7	Add84753 Human alp
22	2293	100.0	427	8	Adj88279 Human wt
23	2293	100.0	427	8	Adm48683 Human wil
24	2293	100.0	427	8	Adu66918 Human alp

25	2293	100.0	427	9	Aea27447 Human alp
26	2293	100.0	429	2	Aar53765 Alpha-gal
27	2293	100.0	429	2	Aar63234 Human alp
28	2293	100.0	429	2	Aar70207 Alpha-gal
29	2293	100.0	429	5	Aae28206 Human wt
30	2293	100.0	429	7	Adf76604 Novel hum
31	2293	100.0	429	8	Adn49740 Human alp
32	2293	100.0	429	8	Adp24703 PRO poly
33	2293	100.0	429	8	Adu18066 Human can
34	2293	100.0	429	8	Adu74416 Human alp
35	2293	100.0	429	9	Ady19459 PRO poly
36	2293	100.0	429	9	Ady17095 PRO poly
37	2293	100.0	429	9	Aea27471 Human alp
38	2293	100.0	429	9	Aeb43188 Human alp
39	2293	100.0	430	7	Abm79001 Human alp
40	2293	100.0	431	5	Aae28209 Human rga
41	2293	100.0	431	7	Add84749 Human alp
42	2293	100.0	431	8	Adj88275 Human wt
43	2293	100.0	431	8	Adm48679 Human wil
44	2293	100.0	431	8	Adu66914 Human alp
45	2293	100.0	431	9	Aea27443 Human alp

ALIGNMENTS

RESULT 1
AAE28212
ID AAE28212 standard; protein; 417 AA.
XX
AC AAE28212;
XX
DT 27-DEC-2002 (first entry)
XX
DE Human rGAL-12 protein.
XX
KW Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
KW therapeutic; rGAL-12.
XX
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT Cleavage-site 401..402
FT /note= "CTPP cleavage site"
XX
PN US2002088024-A1.
XX
PD 04-JUL-2002.
XX
PF 13-NOV-2001; 2001US-00993059.
XX
PR 26-JUL-2000; 2000US-00626127.
XX
PA (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
XX
PI (KUMA/) KUMAGAI M H.
XX
PI Garger SJ, Turpen TH, Kumagai MH;
XX
DR WPI; 2002-681656/73.
DR N-PSDB; AAD45225.
XX
PT Novel human alpha-galactosidase polypeptide useful for treating lysosomal storage diseases.
XX
PS Claim 7; Page 49-50; 88pp; English.
XX
CC The invention relates to human alpha-galactosidase truncated at the carboxy terminus and the production of enzymatically active recombinant human and animal lysosomal enzymes. The invention is useful for producing lysosomal enzymes for treating lysosomal storage diseases, producing altered or mutated proteins, enzymatically active or otherwise, to serve as precursors or substrates for further in vivo or in vitro processing to

CC a specialised industrial form for research or therapeutic uses, to
 CC produce more effective therapeutic enzyme, for producing antibodies
 CC against lysosomal enzymes for medical diagnostic use, and in any
 CC commercial process that involves substrate hydrolysis. The present
 CC sequence is human rGAL-12 protein
 XX
 SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 5; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.3e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
 Db 1 MQLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWAPQDSGRQLQADPQRPFGIRQL 120
 Db 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWAPQDSGRQLQADPQRPFGIRQL 120

Qy 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYDIDAQTFADWGVDLLKFGDGCYCDLSLENL 180
 Db 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYDIDAQTFADWGVDLLKFGDGCYCDLSLENL 180

Qy 181 ADGKMSLALNRTGRSIVVSCWPLYMPPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 Db 181 ADGKMSLALNRTGRSIVVSCWPLYMPPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SIIDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFWNSNDL 300
 Db 241 SIIDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFWNSNDL 300

Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQIEG 360
 Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQIEG 360

Qy 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
 Db 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 2
 ADD84755
 ID ADD84755 standard; protein; 417 AA.

XX
 AC ADD84755;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Human alpha-galactosidase rGAL-12 polypeptide.
 XX
 KW Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
 KW enzyme replacement therapy; lysosomal disease; enzyme.
 XX
 OS Homo sapiens.
 XX
 PN US2003106095-A1.
 XX
 PD 05-JUN-2003.
 XX
 PF 20-MAR-2002; 2002US-00103327.
 XX
 PR 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 XX
 PA (GARG/) GARGER S J.
 PA (TURP/) TURPEN T H.
 PA (KOMA/) KOMAGAI M H.
 XX
 PI Garger SJ, Turpen TH, Kumagai MH;
 XX
 DR WPI; 2003-801257/75.
 DR N-PSDB; ADD84754.

XX New polynucleotide for producing active recombinant human and animal
 PT lysosomal enzymes in a plant expression system that can be used in enzyme
 PT replacement therapy.
 XX
 PS Claim 7; SEQ ID NO 16; 77pp; English.
 XX
 CC The invention relates to human alpha-galactosidase derivatives and the
 CC nucleic acids encoding them. The polypeptides are used in a method for
 CC producing active recombinant human and animal lysosomal enzymes in a
 CC plant expression system. The enzymes can be used in enzyme replacement
 CC therapy for the therapeutic treatment of human and animal lysosomal
 CC diseases. This sequence represents a human alpha-galactosidase derivative
 CC polypeptide of the invention.

SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 7; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.3e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
 Db 1 MQLRNPELHIGCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60

Qy 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWAPQDSGRQLQADPQRPFGIRQL 120
 Db 61 DSCISEKLFMEAEMLVSEGWKDGAYEYLICDDCWAPQDSGRQLQADPQRPFGIRQL 120

Qy 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYDIDAQTFADWGVDLLKFGDGCYCDLSLENL 180
 Db 121 ANYVHSKGLKGIYADVGNKTCAGFPGSFGYDIDAQTFADWGVDLLKFGDGCYCDLSLENL 180

Qy 181 ADGKMSLALNRTGRSIVVSCWPLYMPPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
 Db 181 ADGKMSLALNRTGRSIVVSCWPLYMPPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SIIDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFWNSNDL 300
 Db 241 SIIDWTSFNOERIVDVAGPGGNDPDMLVIGNFGLSNQOVTQMALWAINAAPLFWNSNDL 300

Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQIEG 360
 Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRGDNFEVWERPLSLGLAWAVAMINRQIEG 360

Qy 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
 Db 361 GPRSYTIAVASLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 3

ADJ88281
 ID ADJ88281 standard; protein; 417 AA.

XX
 AC ADJ88281;
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Human WT rGAL-12 (galactosidase).
 XX
 KW Galactosidase; GAL; gene therapy; lysosomal storage disease;
 KW Fabry's disease; Gaucher's disease; human; enzyme.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH Misc-difference 400..417
 FT /note= "Encoded by ACT"
 XX
 PN US2004016021-A1.
 XX
 PD 22-JAN-2004.

```

PF 23-JUN-2003; 2003US-00602219.
XX 26-FEB-1988; 88US-00160766.
PR 26-FEB-1988; 88US-00160771.
PR 15-JUL-1988; 88US-00219279.
PR 17-FEB-1989; 89US-00310881.
PR 05-MAY-1989; 89US-00347637.
PR 08-JUN-1989; 89US-00363138.
PR 22-OCT-1990; 90US-00600244.
PR 16-JAN-1991; 91US-00641617.
PR 26-JUL-1991; 91US-00737899.
PR 01-AUG-1991; 91US-00739143.
PR 31-JUL-1992; 92US-00923692.
PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
PA (POGU/) FOGUE G P.
PA (ERWI/) ERWIN R L.
PA (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LK;
PI
XX WPI; 2004-108227/11.
XX N-PSDB; ADJ88280.
DR
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 7; SEQ ID NO 16; 71pp; English.
PS
XX The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC protein.
XX
XX Sequence 417 AA;
Qy
Query Match 100.0%; Score 2293; DB 8; Length 417;
Best Local Similarity 100.0%; Pred. No. 5.3e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDQDEP 60
Db 1 MOLRNPELHLCALALRFLALVSDIPGARALDNGLARTPTMGWLHWRFCNLDQDEP 60
Qy 61 DSCISEKLFMEALMSEGHKADAGEYELCIDDCWAPORDSEGRLOADPQPFPHGIROL 120
Db 61 DSCISEKLFMEALMSEGHKADAGEYELCIDDCWAPORDSEGRLOADPQPFPHGIROL 120
Qy 121 ANYVHSGKLGIVADVGNKTCAGPFGSGFYDIDAOTFADWGDVLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGIVADVGNKTCAGPFGSGFYDIDAOTFADWGDVLLKFDGCYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVYSCWPLVYWPQKFNTEIRQYCNHRNFADIDDSWSIK 240
Db 181 ADGKMSLALNRTGRSIVYSCWPLVYWPQKFNTEIRQYCNHRNFADIDDSWSIK 240
Qy 241 SILDWTSTNQRIVDVAGPGWNPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Db 241 SILDWTSTNQRIVDVAGPGWNPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Qy 301 RHISFOAKALLQDKVDVAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISFOAKALLQDKVDVAINODPLGKQGYQLRQGNFVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVILLQ 417

```

```

Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGYEWTSLRSHINPTGTVILLQ 417
RESULT 4
ADM48685
ID ADM48685 standard; protein; 417 AA.
XX
XX ADM48685;
XX
XX 03-JUN-2004 (first entry)
XX
XX Human wild type rGAL-12 protein.
DE
XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
KW enzyme.
XX
XX Homo sapiens.
OS
XX
XX US2004023281-A1.
XX
XX 05-FEB-2004.
PD
XX
XX 23-JUN-2003; 2003US-00602220.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 17-FEB-1989; 89US-00310881.
XX 22-OCT-1990; 90US-00600244.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX (POGU/) FOGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
PI
XX
XX WPI; 2004-142650/14.
XX N-PSDB; ADM48684.
XX
XX New alpha-galactosidase polypeptides, useful in producing recombinant
PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
PT disease.
XX
XX Disclosure; SEQ ID NO 16; 72pp; English.
XX
XX The present invention relates to novel galactosidase (Gal) proteins such
XX as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
XX of the present invention are useful for producing recombinant lysosomal
XX enzymes for enzyme replacement therapy for treating human and animal
XX lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
XX disease, Fabry disease and Tay-Sachs disease. The present sequence is
XX human wild type rGAL-12 protein used in the exemplification of the
XX invention.
XX
XX Sequence 417 AA;
SQ
Query Match 100.0%; Score 2293; DB 8; Length 417;
Best Local Similarity 100.0%; Pred. No. 5.3e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


QY 1 MOLNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGHLWHERFMCNLDQCEEP 60
 DB 1 MOLNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGHLWHERFMCNLDQCEEP 60
 QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
 DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
 QY 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 QY 181 ADGKXMSLALNRTGRSIVYSCWPLWYFQKNTYIROYCNHWNFNADIDDSWKSIIK 240
 DB 181 ADGKXMSLALNRTGRSIVYSCWPLWYFQKNTYIROYCNHWNFNADIDDSWKSIIK 240
 QY 241 SILDWTFSNQRERIVDVAGPGGWNDDPMLVIGNFGLSNQVQTQWALWAINAAPLFMSNDL 300
 DB 241 SILDWTFSNQRERIVDVAGPGGWNDDPMLVIGNFGLSNQVQTQWALWAINAAPLFMSNDL 300
 QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLUSGLAWAVAMINRQIEG 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLUSGLAWAVAMINRQIEG 360
 QY 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
 DB 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 5
 ADU66920
 ID ADU66920 standard; protein; 417 AA.
 XX
 AC ADU66920;
 XX
 DT 10-FEB-2005 (first entry)
 XX
 DE Human alpha-galactosidase protein #7.
 XX

XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease;
 KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotropic; human.
 XX
 OS Homo sapiens.

XX US2004234516-A1.
 XX
 XX 25-NOV-2004.
 XX
 XX 21-MAY-2004; 2004US-00851388.
 XX
 XX 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 PR 20-MAR-2002; 2002US-00103327.
 XX
 XX (LARG-) LARGE SCALE BIOLOGY CORP.

XX Garger SJ, Turpen TH, Kumagai MH;
 XX
 XX WPI; 2004-821274/81.
 DR N-PSDB; ADU66919.
 DR

XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
 PT enzyme replacement therapy for the treatment of lysosomal storage
 PT diseases, such as Fabry's disease.
 PT

XX Disclosure; SEQ ID NO 16; 88pp; English.
 XX
 XX The present invention relates to the production of human and animal
 CC lysosomal enzymes in plants by a transient plant expression system. The
 CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
 CC galactosidase enzymes having a post-translational modification provided

CC by the plant expression system. The invention is useful in enzyme
 CC replacement therapy for treating lysosomal storage diseases such as
 CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
 CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
 CC also useful in researches for developing new approaches to medical
 CC treatment of lysosomal storage diseases and in industrial processes
 CC involving enzymatic substrate hydrolysis. The present sequence is the
 CC human alpha-galactosidase protein.

XX Sequence 417 AA;
 SQ

Query Match 100.0%; Score 2293; DB 8; Length 417;
 Best Local Similarity 100.0%; Pred. No. 5.3e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOLNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGHLWHERFMCNLDQCEEP 60
 DB 1 MOLNPELHLCALALRFLALVSWDIPCARALDNGLARTPTMGHLWHERFMCNLDQCEEP 60
 QY 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
 DB 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
 QY 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 DB 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
 QY 181 ADGKXMSLALNRTGRSIVYSCWPLWYFQKNTYIROYCNHWNFNADIDDSWKSIIK 240
 DB 181 ADGKXMSLALNRTGRSIVYSCWPLWYFQKNTYIROYCNHWNFNADIDDSWKSIIK 240
 QY 241 SILDWTFSNQRERIVDVAGPGGWNDDPMLVIGNFGLSNQVQTQWALWAINAAPLFMSNDL 300
 DB 241 SILDWTFSNQRERIVDVAGPGGWNDDPMLVIGNFGLSNQVQTQWALWAINAAPLFMSNDL 300
 QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLUSGLAWAVAMINRQIEG 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLUSGLAWAVAMINRQIEG 360
 QY 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
 DB 361 GPRSYTTIYAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 6
 AEA27449
 ID AEA27449 standard; protein; 417 AA.
 XX
 AC AEA27449;
 XX
 DT 11-AUG-2005 (first entry)
 XX
 DE Human alpha-galactosidase protein, rGAL-12, SEQ ID NO: 16.
 XX
 DE Gauchers disease; metabolic; neurologic disorder; niemann pick disease;
 KW genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
 KW antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
 KW enzyme.
 XX
 XX Homo sapiens.
 OS
 XX US2005125859-A1.
 PN
 XX 09-JUN-2005.
 PD
 XX 08-NOV-2004; 2004US-00984389.
 PF
 XX 26-JUL-2000; 2000US-00626127.
 PR 13-NOV-2001; 2001US-00993059.
 PR 20-MAR-2002; 2002US-00103327.
 XX
 XX (LARG-) LARGE SCALE BIOLOGY CORP.
 XX

```

PI Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
DR N-PSDB; AEA27448.
XX
PT New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
XX Disclosure; SEQ ID NO 16; 88pp; English.
XX
XX The present invention relates to the production of human and animal
XX lysosomal enzymes in plants by a transient plant expression system. The
XX invention relates to glucocerebrosidase (GCB, GCR) and alpha-
XX galactosidase (Gal) enzymes having a post-translational modification
XX provided by the plant expression system. The invention is useful in
XX enzyme replacement therapy for treating lysosomal storage diseases such
XX as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
XX disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
XX also useful in researches for developing new approaches to medical
XX treatment of lysosomal storage diseases and in industrial processes
XX involving enzymatic substrate hydrolysis. The present sequence is the
XX human alpha-galactosidase protein.
SQ Sequence 417 AA;

Query Match 100.0%; Score 2293; DB 9; Length 417;
Best Local Similarity 100.0%; Pred. No. 5.3e-220; Indels 0; Gaps 0;
Matches 417; Conservative 0; Mismatches 0;

QY 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFMCNLDCCQEEP 60
DB 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFMCNLDCCQEEP 60
QY 61 DSCISEKLFMEAMELMVSEGWKDAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAMELMVSEGWKDAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL 180
QY 181 ADGKXMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFNADIDDSWSKIK 240
DB 181 ADGKXMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFNADIDDSWSKIK 240
QY 241 SILDWTSFNQRERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQRERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQIEG 360
DB 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQIEG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTLLQL 417
DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTLLQL 417

RESULT 7
AAE28210
ID AAE28210 standard; protein; 421 AA.
XX
XX AAE28210;
AC
XX
XX 27-DEC-2002 (first entry)
DT
XX
XX Human rGAL-8 protein.
DE
XX
XX Human; alpha-galactosidase; lysosomal enzyme; lysosomal storage disease;
XX therapeutic; rGAL-8.
XX

```

```

OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Cleavage-site 401..402
FT /note="CTPP cleavage site"
XX
XX US2002088024-A1.
PN
XX
XX 04-JUL-2002.
PD
XX
XX 13-NOV-2001; 2001US-00993059.
PF
XX
XX 26-JUL-2000; 2000US-00626127.
PR
XX
XX (GARG/) GARGER S J.
PA (TURP/) TURPEN T H.
PA (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2002-681656/73.
XX N-PSDB; AAD45223.
DR
XX
XX Novel human alpha-galactosidase polypeptide useful for treating lysosomal
XX storage diseases.
PT
XX
XX Claim 7; Page 44; 88pp; English.
XX
XX The invention relates to human alpha-galactosidase truncated at the
XX carboxy terminus and the production of enzymatically active recombinant
XX human and animal lysosomal enzymes. The invention is useful for producing
XX lysosomal enzymes for treating lysosomal storage diseases, producing
XX altered or mutated proteins, enzymatically active or otherwise, to serve
XX as precursors or substrates for further in vivo or in vitro processing to
XX a specialised industrial form for research or therapeutic uses, to
XX produce more effective therapeutic enzyme, for producing antibodies
XX against lysosomal enzymes for medical diagnostic use, and in any
XX commercial process that involves substrate hydrolysis. The present
XX sequence is human rGAL-8 protein
XX
XX Sequence 421 AA;

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFMCNLDCCQEEP 60
DB 1 MQLRNPELHGCALALRFLALVSDIPGARALDNGLARTPTMGWLHWFMCNLDCCQEEP 60
QY 61 DSCISEKLFMEAMELMVSEGWKDAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAMELMVSEGWKDAGYEYLCIDDCWMAQRDSEGRLOADPQRFPHGIRQL 120
QY 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL 180
DB 121 ANYVHSGKLGLGIYADVGNKTCAGFGSGFYDIDAQTFADMGVDLLKFDGCDLSLENL 180
QY 181 ADGKXMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFNADIDDSWSKIK 240
DB 181 ADGKXMSLALNRTGRSIVVSCWPLVWPFQKPNYTEIRQYCNHWNFNADIDDSWSKIK 240
QY 241 SILDWTSFNQRERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIMAAPLFMSNDL 300
DB 241 SILDWTSFNQRERIVDVAGPGGWNDDPMLVGNFGLSNQVQTQWALWAIMAAPLFMSNDL 300
QY 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQIEG 360
DB 301 RHISPOKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQIEG 360
QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTLLQL 417
DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSLRSHINPTGTLLQL 417

```

```
RESULT 8
ADD84751
ID ADD84751 standard; protein; 421 AA.
XX
XX ADD84751;
XX
XX 29-JAN-2004 (first entry)
XX
XX Human alpha-galactosidase rGAL-8 polypeptide.
XX
XX Human; alpha-galactosidase; rGAL-4; lysosomal enzyme;
XX enzyme replacement therapy; lysosomal disease; enzyme.
XX
XX Homo sapiens.
XX
XX US2003106095-A1.
XX
XX 05-JUN-2003.
XX
XX 20-MAR-2002; 2002US-00103327.
XX
XX 26-JUL-2000; 2000US-00626127.
XX
XX 13-NOV-2001; 2001US-00993059.
XX
XX (GARG/) GARGER S J.
XX (TURP/) TURPEN T H.
XX (KUMA/) KUMAGAI M H.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX
XX WPI; 2003-801257/75.
XX N-PSDB; ADD84750.
XX
XX New polynucleotide for producing active recombinant human and animal
XX lysosomal enzymes in a plant expression system that can be used in enzyme
XX replacement therapy.
XX
XX Claim 7; SEQ ID NO 12; 77pp; English.
XX
XX The invention relates to human alpha-galactosidase derivatives and the
XX nucleic acids encoding them. The polypeptides are used in a method for
XX producing active recombinant human and animal lysosomal enzymes in a
XX plant expression system. The enzymes can be used in enzyme replacement
XX therapy for the therapeutic treatment of human and animal lysosomal
XX diseases. This sequence represents a human alpha-galactosidase derivative
XX polypeptide of the invention.
XX
XX Sequence 421 AA;
XX
Query Match 100.0%; Score 2293; DB 7; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWFMCNLDCEEP 60
Dy 1 MOLNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWFMCNLDCEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKQAGYEYLCIDDCWMAPOQDSEGRQLQADPQRPFGIRQL 120
Dy 61 DSCISEKLFMEAEMLVSEGWKQAGYEYLCIDDCWMAPOQDSEGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSPGYDIDAQTFADMGVDLLKFDGCYCDLSLENL 180
Dy 121 ANYVHSGKLGLGIYADVGNKTCAGFPGSPGYDIDAQTFADMGVDLLKFDGCYCDLSLENL 180
Qy 181 ADGKYHMSLALNRGRSIVVSCWPLVWMPQKENYTEIRQYCNHWRNFADIDDSWKSIIK 240
Dy 181 ADGKYHMSLALNRGRSIVVSCWPLVWMPQKENYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOQVTOQWALWMAAPLFMSNDL 300
Dy 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOQVTOQWALWMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOQVTOQWALWMAAPLFMSNDL 300
Qy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWVERPLSGLAWAVAMINROEIG 360
Dy 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWVERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOAKALLQDKDVIAINQDPLGKQGYOLROGDNFVWVERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Dy 361 GPRSYTIAVASLGLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLGLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
RESULT 9
ADJ88277
ID ADJ88277 standard; protein; 421 AA.
XX
XX ADJ88277;
XX
XX 06-MAY-2004 (first entry)
XX
XX Human WT rGAL-8 (galactosidase).
XX
XX Galactosidase; GAL; gene therapy; lysosomal storage disease;
XX Fabry's disease; Gaucher's disease; human; enzyme.
XX
XX Homo sapiens.
XX
XX US2004016021-A1.
XX
XX 22-JAN-2004.
XX
XX 23-JUN-2003; 2003US-00602219.
XX
XX 26-FEB-1988; 88US-00160766.
XX 26-FEB-1988; 88US-00160771.
XX 15-JUL-1988; 88US-00219279.
XX 17-FEB-1989; 89US-00310881.
XX 05-MAY-1989; 89US-00347637.
XX 08-JUN-1989; 89US-00363138.
XX 22-OCT-1990; 90US-00600244.
XX 16-JAN-1991; 91US-00641617.
XX 26-JUL-1991; 91US-00737899.
XX 01-AUG-1991; 91US-00739143.
XX 31-JUL-1992; 92US-00923692.
XX 30-DEC-1992; 92US-00997733.
XX 29-DEC-1993; 93US-00176414.
XX 19-JAN-1994; 94US-00184237.
XX 14-OCT-1994; 94US-00324003.
XX 21-MAY-1999; 99US-00316572.
XX 26-JUL-2000; 2000US-00626127.
XX 13-NOV-2001; 2001US-00993059.
XX
XX (TURP/) TURPEN T H.
XX (FOGU/) POGUE G P.
XX (ERWI/) ERWIN R L.
XX (GRIL/) GRILL L K.
XX
XX Turpen TH, Pogue GP, Erwin RL, Grill LX;
XX
XX WPI; 2004-108227/11.
XX N-PSDB; ADJ88276.
XX
XX New lysosomal enzymes, useful in treating human and animal lysosomal
XX storage diseases, e.g. Fabry's disease and Gaucher's diseases.
XX
XX Claim 7; SEQ ID NO 12; 71pp; English.
XX
XX The invention relates to nucleotide encoding galactosidase (GAL). The
XX invention is useful in gene therapy. The polynucleotides and polypeptides
XX are useful in treating human and animal lysosomal storage diseases, e.g.
XX Fabry's disease and Gaucher's diseases. The present sequence is human GAL
XX protein.
XX
XX Sequence 421 AA;
XX
```

Query Match 100.0%; Score 2293; DB 8; Length 421;
 Best Local Similarity 100.0%; Pred. No. 5.4e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
 DB 1 MQLRNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60

QY 61 DSCISEKLFMEMAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQORFPHGIRQL 120
 DB 61 DSCISEKLFMEMAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQORFPHGIRQL 120

QY 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
 DB 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSEFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
 DB 241 SILDWTSEFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360

QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417
 DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417

RESULT 10
 ADW48681
 ID ADW48681 standard; protein; 421 AA.
 AC ADW48681;
 DT 03-JUN-2004 (first entry)
 XX Human wild type rGAL-8 protein.
 XX Galactosidase; Gal; lysosomal enzyme; enzyme replacement therapy;
 KW lysosomal storage disease; Gaucher's disease; Niemann-Pick disease;
 KW Fabry disease; Tay-Sachs disease; cardiovascular; nephrotrophic; human;
 KW enzyme.
 XX Homo sapiens.
 XX US2004023281-A1.
 XX 05-FEB-2004.
 XX 23-JUN-2003; 2003US-00602220.
 XX 26-FEB-1988; 88US-00160766.
 XX 26-FEB-1988; 88US-00160771.
 XX 17-FEB-1989; 89US-00310881.
 XX 22-OCT-1990; 90US-00600244.
 XX 31-JUL-1992; 92US-00923692.
 XX 30-DEC-1992; 92US-00997733.
 XX 29-DEC-1993; 93US-00176414.
 XX 19-JAN-1994; 94US-00184237.
 XX 14-OCT-1994; 94US-00324003.
 XX 21-MAY-1999; 99US-00316572.
 XX 26-JUL-2000; 2000US-00626127.
 XX 13-NOV-2001; 2001US-00993059.
 XX (TURP/) TURPEN T H.
 PA (KUMA/) KUMAGAI M H.
 PA (POGU/) POGUE G P.

PA (ERWI/) ERWIN R L.
 PA (GRIL/) GRILL L K.
 PI Turpen TH, Kumagai MH, Pogue GP, Erwin RL, Grill LK;
 XX WPI; 2004-142650/14.
 DR N-PSDB; ADM48680.
 XX New alpha-galactosidase polypeptides, useful in producing recombinant
 PT lysosomal enzymes for the treatment of lysosomal storage diseases, such
 PT as Gaucher's disease, Niemann-Pick disease, Fabry disease and Tay-Sachs
 PT disease.
 XX Disclosure; SEQ ID NO 12; 72pp; English.
 XX The present invention relates to novel galactosidase (Gal) proteins such
 CC as rGAL-12, rGAL-12R, r-GAL-25 or rGAL-25R. The methods and compositions
 CC of the present invention are useful for producing recombinant lysosomal
 CC enzymes for enzyme replacement therapy for treating human and animal
 CC lysosomal storage diseases such as Gaucher's disease, Niemann-Pick
 CC disease, Fabry disease and Tay-Sachs disease. The present sequence is
 CC human wild type rGAL-8 protein used in the exemplification of the
 CC invention.
 XX Sequence 421 AA;
 SQ

Query Match 100.0%; Score 2293; DB 8; Length 421;
 Best Local Similarity 100.0%; Pred. No. 5.4e-220;
 Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60
 DB 1 MQLRNPELHGCALALRFLALVSDIFGARALDNGLARTPTMGWLHWRFCNLDCCOEP 60

QY 61 DSCISEKLFMEMAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQORFPHGIRQL 120
 DB 61 DSCISEKLFMEMAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADQORFPHGIRQL 120

QY 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
 DB 121 ANYVHSGKLGKIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
 DB 181 ADGKMSLALNRTGRSIVYSCWPLYMWPFOKPNYTEIRQYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSEFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
 DB 241 SILDWTSEFQERIVDVAGPGGNDPDMVLVGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360
 DB 301 RHISPOKALLQDKVDIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQIEG 360

QY 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417
 DB 361 GPRSYTIAVASLKGKGVACNCPACFITQLLPVKRKLGFYEWTSRLSRSHINPTGTLLQL 417

RESULT 11
 ADU66916
 ID ADU66916 standard; protein; 421 AA.
 XX ADU66916;
 XX 10-FEB-2005 (first entry)
 DE Human alpha-galactosidase protein #5.
 XX Lysosomal enzyme; glucocerebrosidase; GCB; GCR; alpha-galactosidase;
 KW enzyme replacement therapy; lysosomal storage disease; Gaucher's disease;
 KW Niemann-Pick disease; Fabry's disease; Tay-Sachs disease; Hurler's
 KW Hurler's syndrome; Hurler-Scheie syndrome; nephrotrophic; human.

```
XX Homo sapiens.
OS
XX AEA27445
PN US2004234516-A1.
XX
XX 25-NOV-2004.
XX
XX 21-MAY-2004; 2004US-00851388.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2004-821274/81.
XX N-PSDB; ADU66915.
XX
XX A pharmaceutical composition comprising a lysosomal enzyme, useful for
PT enzyme replacement therapy for the treatment of lysosomal storage
PT diseases, such as Fabry's disease.
XX
XX Disclosure; SEQ ID NO 12; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase enzymes having a post-translational modification provided
CC by the plant expression system. The invention is useful in enzyme
CC replacement therapy for treating lysosomal storage diseases such as
CC Gaucher's disease, Niemann-Pick disease, Fabry's disease and Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
XX Sequence 421 AA;
XX
Query Match 100.0%; Score 2293; DB 8; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPEHLHGICALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
Db 1 MQLRNPEHLHGICALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAGVEYLCIDDCWMAQRDSEGRLOADQRPFGHGRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAGVEYLCIDDCWMAQRDSEGRLOADQRPFGHGRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Qy 181 ADGKMSLALNTRGRISVSCWPLFMWPFQPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Db 181 ADGKMSLALNTRGRISVSCWPLFMWPFQPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
Qy 241 SILDWTSFNQRIVDVAGPGGNDPDMLVINGFGLSNQOVTOMALWAIWAAPLFMSNDL 300
Db 241 SILDWTSFNQRIVDVAGPGGNDPDMLVINGFGLSNQOVTOMALWAIWAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLISGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLISGLAWAVAMINRQEI 360
Qy 361 GPRSYTTIAVSLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTTIAVSLGKGVACNPACFTITQLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417
```

```
RESULT 12
AEA27445
ID AEA27445 standard; protein; 421 AA.
XX
XX AC AEA27445;
XX
XX 11-AUG-2005 (first entry)
XX
XX Human alpha-galactosidase protein, rGAL-8, SEQ ID NO: 12.
XX
XX Gauchers disease; metabolic; neurological disease; niemann pick disease;
XX genetic disorder; Fabry disease; metabolic disorder; tay sachs disease;
XX antilipemic; cns-gen.; lysosome storage disease; alpha-galactosidase;
XX enzyme.
XX
XX Homo sapiens.
XX
XX US2005125859-A1.
XX
XX 09-JUN-2005.
XX
XX 08-NOV-2004; 2004US-00984389.
XX
XX 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
PR 20-MAR-2002; 2002US-00103327.
XX
XX (LARG-) LARGE SCALE BIOLOGY CORP.
XX
XX Garger SJ, Turpen TH, Kumagai MH;
XX WPI; 2005-404004/41.
XX N-PSDB; AEA27444.
XX
XX New isolated polypeptides useful for producing lysosomal enzymes in
PT plants to be utilized in enzyme replacement therapy or for the
PT therapeutic treatment of human or animal lysosomal storage diseases, e.g.
PT Gaucher's disease.
XX
XX Claim 28; SEQ ID NO 12; 88pp; English.
XX
XX The present invention relates to the production of human and animal
CC lysosomal enzymes in plants by a transient plant expression system. The
CC invention relates to glucocerebrosidase (GCB, GCR) and alpha-
CC galactosidase (Gal) enzymes having a post-translational modification
CC provided by the plant expression system. The invention is useful in
CC enzyme replacement therapy for treating lysosomal storage diseases such
CC as Gaucher's disease, Niemann-Pick disease, Fabry's disease, Tay-Sachs
CC disease, Hurler's syndrome and Hurler-Scheie syndrome. The invention is
CC also useful in researches for developing new approaches to medical
CC treatment of lysosomal storage diseases and in industrial processes
CC involving enzymatic substrate hydrolysis. The present sequence is the
CC human alpha-galactosidase protein.
XX
XX Sequence 421 AA;
XX
Query Match 100.0%; Score 2293; DB 9; Length 421;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLRNPEHLHGICALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
Db 1 MQLRNPEHLHGICALALRFLALVSWDIPGARALDNLGLARTMTGWLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKDAGVEYLCIDDCWMAQRDSEGRLOADQRPFGHGRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKDAGVEYLCIDDCWMAQRDSEGRLOADQRPFGHGRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWVDLLKFDGCYCDLSLENL 180
```


CC producing active recombinant human and animal lysosomal enzymes in a
CC plant expression system. The enzymes can be used in enzyme replacement
CC therapy for the therapeutic treatment of human and animal lysosomal
CC diseases. This sequence represents a human alpha-galactosidase derivative
CC polypeptide of the invention.
XX
SQ Sequence 423 AA;

Query Match 100.0%; Score 2293; DB 7; Length 423;
Best Local Similarity 100.0%; Pred. No. 5.4e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDGLARTPTMGWLHWRFCMLDCCQEEP 60
DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDGLARTPTMGWLHWRFCMLDCCQEEP 60
QY 61 DSCISEKLFMEAEMLVSEGWKADGVEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADGVEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
QY 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVLLKFDGVCYDLSLENL 180
DB 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVLLKFDGVCYDLSLENL 180
QY 181 ADGKMSLALNRTGRSIVVSCWPLYMFPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
DB 181 ADGKMSLALNRTGRSIVVSCWPLYMFPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240
QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
DB 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300
QY 301 RHISPOKALLQDKVDVAINQDPLGKQYQIQDGFVWERPLSGLAWAVAMINRQIEG 360
DB 301 RHISPOKALLQDKVDVAINQDPLGKQYQIQDGFVWERPLSGLAWAVAMINRQIEG 360
QY 361 GPRSYTIAVSLGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTIAVSLGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 15

ADJ98283

ID ADJ98283 standard; protein; 423 AA.

XX AC ADJ98283;

XX 06-MAY-2004 (first entry)

XX Human WT rGAL-12R (galactosidase).

XX Galactosidase; GAL; gene therapy; lysosomal storage disease;

XX Fabry's disease; Gaucher's disease; human; enzyme.

XX OS Homo sapiens.

XX PN US2004016021-A1.

XX 22-JAN-2004.

XX 23-JUN-2003; 2003US-00602219.

XX 26-FEB-1988; 88US-00160766.

XX 26-FEB-1988; 88US-00160771.

XX 15-JUL-1988; 88US-00219279.

XX 17-FEB-1989; 89US-00310881.

XX 05-MAY-1989; 89US-00347637.

XX 08-JUN-1989; 89US-00363138.

XX 22-OCT-1990; 90US-00600244.

XX 16-JAN-1991; 91US-00641617.

XX 26-JUL-1991; 91US-00737899.

XX 01-AUG-1991; 91US-00739143.

XX 31-JUL-1992; 92US-00923692.

PR 30-DEC-1992; 92US-00997733.
PR 29-DEC-1993; 93US-00176414.
PR 19-JAN-1994; 94US-00184237.
PR 14-OCT-1994; 94US-00324003.
PR 21-MAY-1999; 99US-00316572.
PR 26-JUL-2000; 2000US-00626127.
PR 13-NOV-2001; 2001US-00993059.
XX

(TURP/) TURPEN T H.
(FOGU/) FOGUE G P.
(ERWI/) ERWIN R L.
(GRIL/) GRILL L K.

Turpen TH, Pogue GP, Erwin RL, Grill LK;

WPI; 2004-108227/11.

N-PSDB; ADJ98282.

XX New lysosomal enzymes, useful in treating human and animal lysosomal
PT storage diseases, e.g. Fabry's disease and Gaucher's diseases.

XX Claim 7; SEQ ID NO 18; 71pp; English.

XX The invention relates to nucleotide encoding galactosidase (GAL). The
CC invention is useful in gene therapy. The polynucleotides and polypeptides
CC are useful in treating human and animal lysosomal storage diseases, e.g.
CC Fabry's disease and Gaucher's diseases. The present sequence is human GAL
CC protein.

XX SQ Sequence 423 AA;

Query Match 100.0%; Score 2293; DB 8; Length 423;

Best Local Similarity 100.0%; Pred. No. 5.4e-220;

Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDGLARTPTMGWLHWRFCMLDCCQEEP 60

DB 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDGLARTPTMGWLHWRFCMLDCCQEEP 60

QY 61 DSCISEKLFMEAEMLVSEGWKADGVEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

DB 61 DSCISEKLFMEAEMLVSEGWKADGVEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

QY 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVLLKFDGVCYDLSLENL 180

DB 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVLLKFDGVCYDLSLENL 180

QY 181 ADGKMSLALNRTGRSIVVSCWPLYMFPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240

DB 181 ADGKMSLALNRTGRSIVVSCWPLYMFPQKPNYTEIRQYCNHWNFNADIDDSWKSIIK 240

QY 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300

DB 241 SILDWTSTFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQTQWALWAIWAAPLFMSNDL 300

QY 301 RHISPOKALLQDKVDVAINQDPLGKQYQIQDGFVWERPLSGLAWAVAMINRQIEG 360

DB 301 RHISPOKALLQDKVDVAINQDPLGKQYQIQDGFVWERPLSGLAWAVAMINRQIEG 360

QY 361 GPRSYTIAVSLGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

DB 361 GPRSYTIAVSLGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

Search completed: December 31, 2005, 23:43:14

Job time : 155.421 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:36:26 ; Search time 43.1379 Seconds
(without alignments)
799.198 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MOLRNPELHGCALALRFLA.....EWTSLRSHINPTGTLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2293	100.0	417	2	US-09-993-059-16
2	2293	100.0	417	2	US-10-103-327-16
3	2293	100.0	421	2	US-09-993-059-12
4	2293	100.0	421	2	US-10-103-327-12
5	2293	100.0	423	2	US-09-993-059-18
6	2293	100.0	423	2	US-10-103-327-18
7	2293	100.0	427	2	US-09-993-059-14
8	2293	100.0	427	2	US-10-103-327-14
9	2293	100.0	429	1	US-07-602-824-2
10	2293	100.0	429	1	US-07-602-608-3
11	2293	100.0	429	1	US-07-983-451-2
12	2293	100.0	429	1	US-08-261-578-3
13	2293	100.0	429	1	US-08-261-577-7
14	2293	100.0	429	1	US-08-261-577-9
15	2293	100.0	431	2	US-09-993-059-10
16	2293	100.0	431	2	US-10-103-327-10
17	2293	100.0	435	2	US-09-993-059-6
18	2293	100.0	435	2	US-10-103-327-6
19	2289	99.8	429	2	US-09-070-356-4
20	2288	99.8	424	2	US-09-993-059-8
21	2288	99.8	424	2	US-10-103-327-8
22	2288	99.8	428	2	US-09-993-059-4
23	2288	99.8	428	2	US-10-103-327-4
24	2275	99.2	428	6	5179023-4
25	2226	97.1	409	2	US-09-993-059-22
26	2226	97.1	409	2	US-10-103-327-22
27	2213	96.5	401	2	US-09-993-059-20

28	2213	96.5	401	2	US-10-103-327-20	Sequence 20, Appl
29	2131	92.9	386	2	US-09-176-666-11	Sequence 11, Appl
30	2131	92.9	387	2	US-09-176-666-10	Sequence 10, Appl
31	2131	92.9	388	2	US-09-176-666-9	Sequence 9, Appl
32	2131	92.9	389	2	US-09-176-666-8	Sequence 8, Appl
33	2131	92.9	390	2	US-09-176-666-7	Sequence 7, Appl
34	2131	92.9	391	2	US-09-176-666-6	Sequence 6, Appl
35	2131	92.9	392	2	US-09-176-666-5	Sequence 5, Appl
36	2131	92.9	393	2	US-09-176-666-4	Sequence 4, Appl
37	2131	92.9	394	2	US-09-176-666-3	Sequence 3, Appl
38	2131	92.9	396	2	US-09-176-666-2	Sequence 2, Appl
39	2131	92.9	398	2	US-08-928-881-26	Sequence 26, Appl
40	2131	92.9	398	2	US-09-176-666-1	Sequence 1, Appl
41	2131	92.9	398	2	US-09-543-921-26	Sequence 26, Appl
42	2131	92.9	398	2	US-09-266-014-4	Sequence 26, Appl
43	2131	92.9	398	2	US-09-491-759-26	Sequence 26, Appl
44	2131	92.9	398	2	US-10-360-101-202	Sequence 202, App
45	2110	92.0	381	2	US-09-176-666-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-993-059-16
; Sequence 16, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monco H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-16

Query Match	100.0%;	Score 2293;	DB 2;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 1.5e-232;		
Matches	417;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	MOLRNPELHGCALALRFLA	VS	WDIPGALDNLGLARTPTMGWLHWRFCNLDCCBEP 60
Db	1	MOLRNPELHGCALALRFLA	VS	WDIPGALDNLGLARTPTMGWLHWRFCNLDCCBEP 60
Qy	61	DSCTSEKLFMEALMVSEGWKDGAYELCTDCCWAPQDSEGRLOADPQRFPHGIQOL 120		
Db	61	DSCTSEKLFMEALMVSEGWKDGAYELCTDCCWAPQDSEGRLOADPQRFPHGIQOL 120		
Qy	121	ANYVHSGKLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWYDLLKFDGCDLSLENL 180		
Db	121	ANYVHSGKLGIYADVGNKTCAGFGPSFGYDIDAQTFADGWYDLLKFDGCDLSLENL 180		
Qy	181	ADGKMSLALNRGSRISVYSCWPLYMFWFKPNYTEIRQYCNHWRNFADIDSWSKIK 240		
Db	181	ADGKMSLALNRGSRISVYSCWPLYMFWFKPNYTEIRQYCNHWRNFADIDSWSKIK 240		
Qy	241	SILDWTSFNQERIVDVAGPGWNDPDMVLVGNFGLSNQOVTQWALWAIMAALPMSNDL 300		
Db	241	SILDWTSFNQERIVDVAGPGWNDPDMVLVGNFGLSNQOVTQWALWAIMAALPMSNDL 300		
Qy	301	RHISPAKALLQDKVIAINQDPLGKQYQLRGQDNFVWVERPLSGLAWAMINRQSIG 360		
Db	301	RHISPAKALLQDKVIAINQDPLGKQYQLRGQDNFVWVERPLSGLAWAMINRQSIG 360		
Qy	361	GPRSYTIAVSLGKGVACNPACFITQLLLPVKRKLGFYEWTSRLRSHINPTGTLLQL 417		

```
|||||
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 2
US-10-103-327-16
; Sequence 16, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/10/103,327
; FILE REFERENCE: 008010087CPUS06
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match 100.0%; Score 2293; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.5e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Qy 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 3
US-09-993-059-12
; Sequence 12, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/09/993,059
; FILE REFERENCE: 008010087CPUS06
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2293; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Qy 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417

RESULT 4
US-10-103-327-12
; Sequence 12, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; CURRENT APPLICATION NUMBER: US/10/103,327
; FILE REFERENCE: 008010087CPUS06
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match 100.0%; Score 2293; DB 2; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Db 1 MQLRNPELHLCALALRFLALVSWDIPGARALDNGLARTPTMGLHWRFCNLDCEEP 60
Qy 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Db 61 DSCISEKLFMEALMVSEGWKADAGEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Db 121 ANYVHSGKLGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGCDLLENL 180
Qy 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGYKMSLALNRTGRSIVSCWPLYMFPQKPNTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNOERIVDVAGPGWMDPDMVLVGNFGLSNQOVTQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYLROGDNFEVWERPLSLGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRLGFYEWTSRLRSHINPTGTLLQL 417
```

Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 5

US-09-993-059-18
; Sequence 18, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-18

Query Match 100.0%; Score 2293; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Db 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360

Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 6

US-10-103-327-18
; Sequence 18, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match 100.0%; Score 2293; DB 2; Length 423;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Db 1 MOLRNPDLHLCALALRFLALVSDIPGARALDNGLARTPTMGLHWRFCMLDCQEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADGAYEYLCIDDCWMAPOQDSGRQLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGKGIYADVGNKTCAGFPQSGFYDDIDQTFADWGVVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGCKMSLALNRGTSIVVSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Qy 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVTQWALWAIMAALFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQYQLOGDNFEVWERPLSGLAWAVAMINRQEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 7

US-09-993-059-14
; Sequence 14, Application US/09993059
; Patent No. 6887696
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPS06
; CURRENT APPLICATION NUMBER: US/09/993,059

```
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-14

Query Match      100.0%; Score 2293; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
Db 1 MQLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60

QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAPOQDSGRQLQADPQRPFGHGIQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAPOQDSGRQLQADPQRPFGHGIQL 120

QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVOTOMALWAIWAAPLFWNSDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVOTOMALWAIWAAPLFWNSDL 300

QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 8
US-10-103-327-14
; Sequence 14, Application US/10103327
; Patent No. 6890748
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 427
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-14

Query Match      100.0%; Score 2293; DB 2; Length 427;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
Db 1 MQLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60

QY 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAPOQDSGRQLQADPQRPFGHGIQL 120
Db 61 DSCISEKLFMEAEMLVSEGKWDAGYEYLCIDCWMAPOQDSGRQLQADPQRPFGHGIQL 120

QY 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGYGYDIDAQTFADMGVDLLKFDGCGYCDLSLENL 180

QY 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240
Db 181 ADGYKMSLALNRTGRSIVYSCWPLYMPPFQKPNYTEIRQYCNHWRNFADIDDSWKSIX 240

QY 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVOTOMALWAIWAAPLFWNSDL 300
Db 241 SILDWTSFQERIVDVAGPGGNDPDMVLVGNFGLSNQOQVOTOMALWAIWAAPLFWNSDL 300

QY 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEI 360

QY 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 9
US-07-602-824A-2
; Sequence 2, Application US/07602824A
; Patent No. 5356804
; GENERAL INFORMATION:
; APPLICANT: Deanick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/602,824A
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-602-824A-2

Query Match      100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPHELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRPFCMLDCQEEP 60
```

Db 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAQDSEGRLOADQPPPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAQDSEGRLOADQPPPHGIRQL 120
Qy 121 ANYVHSGKLGIIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Db 121 ANYVHSGKLGIIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVYSCWPLVMPFQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRTGRSIVYSCWPLVMPFQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFEVWERPLSLGLAWAVAMINRQEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFEVWERPLSLGLAWAVAMINRQEG 360
Qy 361 GPRSYYTAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQL 417
Db 361 GPRSYYTAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQL 417

RESULT 10
US-07-602-608-3
; Sequence 3, Application US/07602608
; Patent No. 5382524
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYL GALACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/07/602,608
; FILING DATE: 24-OCT-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-602-608-3

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
Db 1 MOLNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCNLDQCEEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAQDSEGRLOADQPPPHGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKAGYEYLCIDDCWMAQDSEGRLOADQPPPHGIRQL 120
Qy 121 ANYVHSGKLGIIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Db 121 ANYVHSGKLGIIYADVGNKTCAGPGSGFYDIDAQTFADGWVDLLKFGDCYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVYSCWPLVMPFQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRTGRSIVYSCWPLVMPFQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Db 241 SILDTSFNQERIVDVAGPGGWNDDMLVIGNFGLSNQVQTMALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFEVWERPLSLGLAWAVAMINRQEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQGDNFEVWERPLSLGLAWAVAMINRQEG 360
Qy 361 GPRSYYTAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQL 417
Db 361 GPRSYYTAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSRSHINPTGTVLQL 417

RESULT 11
US-07-983-451-2
; Sequence 2, Application US/07983451
; Patent No. 5401650
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/983,451
; FILING DATE: 30-NOV-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 07/983,451
; REFERENCE/DOCKET NUMBER: 6923-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
; US-07-983-451-2
;
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-261-578-3
;
Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 1 MQLRNPDLHGLCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRERFCNLDCCQEEP 60
DB 1 MQLRNPDLHGLCALALRFLALVSWDIPGARALDNLGLARTPTMGMLHWRERFCNLDCCQEEP 60
;
QY 61 DSCISKLFMEAEMLVSEGWKADGVEYLCIDDCWMAQRDSEGRQLADPQRFPHGIRQL 120
DB 61 DSCISEKLFMEAEMLVSEGWKADGVEYLCIDDCWMAQRDSEGRQLADPQRFPHGIRQL 120
;
QY 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENT 180
DB 121 ANYVHSGKLGIGIYADVGNKTCAGPFGSGYGYDIDAQTFADGWVDLLKFDGCYCDLSLENT 180
;
QY 181 ADGYKMSLALNRTGRSIVYSCWPLVWMPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
DB 181 ADGYKMSLALNRTGRSIVYSCWPLVWMPFQKPNYTEIROYCNHWRNFADIDDSWKSIIK 240
;
QY 241 SILDWTFSFQRIYDVAGPGWNDPDLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
DB 241 SILDWTFSFQRIYDVAGPGWNDPDLVIGNFGLSNQOVTQMALWAIMAAPLFMSNDL 300
;
QY 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINROEIG 360
DB 301 RHISPOAKALLQDKVIAINQDPLGKQGYQLRQGNFVWERPLSGLAWAVAMINROEIG 360
;
QY 361 GPRSYTIAVASLGKGVACNPACFIQTLLPVKRLKGFYEWTSRLRSHINPTGTVLQL 417
DB 361 GPRSYTIAVASLGKGVACNPACFIQTLLPVKRLKGFYEWTSRLRSHINPTGTVLQL 417
;
RESULT 12
US-08-261-578-3
; Sequence 3, Application US/08261578
; Patent No. 5491075
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Wang, Anne M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
; TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYLGLACTOSAMINIDASE
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,578
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 07/602,608
; FILING DATE: 24-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; NAME: Coruzzi, Laura A.
; ATTORNEY/AGENT INFORMATION:
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US/08/261,577
; CURRENT APPLICATION DATA:
; SOFTWARE: Patent In Release #1.0, Version #1.25
; OPERATING SYSTEM: PC-DOS/MS-DOS
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; ZIP: 10036
; STATE: New York
; CITY: New York
; STREET: 1155 Avenue of the Americas
; ADDRESSEE: Pennie & Edmonds
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 12
; TITLE OF INVENTION: Active alpha-Galactosidase A
; TITLE OF INVENTION: Cloning and Expression of Biologically
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Bishop, David F.
; APPLICANT: Desnick, Robert J.
; GENERAL INFORMATION:
; Patent No. 5580757
; Sequence 7, Application US/08261577
; US-08-261-577-7
;
RESULT 13
US-08-261-577-7
; Sequence 7, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; TELEPHONE: (212) 790-9090
; NAME: Coruzzi, Laura A.
; ATTORNEY/AGENT INFORMATION:
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US/08/261,577
; CURRENT APPLICATION DATA:
; SOFTWARE: Patent In Release #1.0, Version #1.25
; OPERATING SYSTEM: PC-DOS/MS-DOS
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
; ZIP: 10036
; STATE: New York
; CITY: New York
; STREET: 1155 Avenue of the Americas
; ADDRESSEE: Pennie & Edmonds
; CORRESPONDENCE ADDRESS:
; NUMBER OF SEQUENCES: 12
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; APPLICANT: Ioannou, Yiannis A.
; APPLICANT: Bishop, David F.
; APPLICANT: Desnick, Robert J.
; GENERAL INFORMATION:
; Patent No. 5580757
; Sequence 7, Application US/08261577
; US-08-261-577-7
;

```

```

; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-261-577-7

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLRNPELHLCALALRFLALVSWDIFGARALDNGLARTPTMGHLHWRFCNLDCEEP 60
DB 1 MQLRNPELHLCALALRFLALVSWDIFGARALDNGLARTPTMGHLHWRFCNLDCEEP 60
QY 61 DSCISEKLFMEABLMVSEGWKADAGBYELCTDDCWMAPOQDSEGRLOADPORFPHGIRQL 120
DB 61 DSCISEKLFMEABLMVSEGWKADAGBYELCTDDCWMAPOQDSEGRLOADPORFPHGIRQL 120
QY 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWGVDLLKFGCYCDSL 180
DB 121 ANYVHSGKGLGIYADVGNKTCAGFPGSFGYDIDAQTFADGWGVDLLKFGCYCDSL 180
QY 181 ADGYKMSLALNRTGRSIVTSCWPLVMWPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
DB 181 ADGYKMSLALNRTGRSIVTSCWPLVMWPFQKPNYTEIRQYCNHWRNFADIDDSWKS 240
QY 241 SILDWTSENOERIVDVAGPGGWNDDPMLVTGNGLSNQVQVQWALWAINAAPLFMSNDL 300
DB 241 SILDWTSENOERIVDVAGPGGWNDDPMLVTGNGLSNQVQVQWALWAINAAPLFMSNDL 300
QY 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQ 360
DB 301 RHISPOAKALLQDKDVIAINQDPLGKQGYQLRGQDNFEVWERPLSLGLAWAVAMINRQ 360
QY 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417
DB 361 GPRSYTIAVASLKGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417

RESULT 14
US-08-261-577-9
; Sequence 9, Application US/08261577
; Patent No. 5580757
; GENERAL INFORMATION:
; APPLICANT: Desnick, Robert J.
; APPLICANT: Bishop, David F.
; APPLICANT: Ioannou, Yiannis A.
; TITLE OF INVENTION: Cloning and Expression of Biologically
; TITLE OF INVENTION: Active alpha-Galactosidase A
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,577
; FILING DATE: 17-JUN-1994

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 6923-042
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-261-577-9

Query Match 100.0%; Score 2293; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPFLHGCALALRFLALYSWDIPGARALDNLGLARTPTMGWLHWRFCMCLDCQEEP 60
Db 1 MQLRNPFLHGCALALRFLALYSWDIPGARALDNLGLARTPTMGWLHWRFCMCLDCQEEP 60
Qy 61 DSCISEKLFMEAEALMVSEGWKDAQGYEYL CIDCWMAPORDSEGRLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEALMVSEGWKDAQGYEYL CIDCWMAPORDSEGRLQADPQRPFGIRQL 120
Qy 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGYDYDIDAQTADMGVDLLKFDGGCYCDSLENL 180
Db 121 ANYVHSKGLKGIYADVGNKTCAGPFGSGYDYDIDAQTADMGVDLLKFDGGCYCDSLENL 180
Qy 181 ADGQKMSALANRTGRSIVYSCWPLYMPPFKPNYTEIRQCNHWRNFADIDDSWKS IK 240
Db 181 ADGQKMSALANRTGRSIVYSCWPLYMPPFKPNYTEIRQCNHWRNFADIDDSWKS IK 240
Qy 241 SILDWTSTFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTQMALWAIAAPL FMSNDL 300
Db 241 SILDWTSTFNOERIVDVAGPGGWNDDPMLVIGNFGLSWNQOVTQMALWAIAAPL FMSNDL 300
Qy 301 RHISPOAKALLQDKQVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEIG 360
Db 301 RHISPOAKALLQDKQVIAINQDPLGKQGYQLRGQDNFVWERPLSLGLAWAVAMINRQEIG 360
Qy 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417
Db 361 GPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTVLLQL 417

RESULT 15
US-09-993-059-10
; Sequence 10, Application US/09993059
; Patent No. 688766
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KIMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-10

Query Match 100.0%; Score 2293; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.6e-232;

```


Matches 417;		Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MQLRNPFLHLCALALRFLALVSWDIFGARALDNGLARTPTMGWLHWRPWCNLDCEEP 60							
Db	1	MQLRNPFLHLCALALRFLALVSWDIFGARALDNGLARTPTMGWLHWRPWCNLDCEEP 60							
Qy	61	DSCISEKLFMEMAELMVSEGWKDGAYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120							
Db	61	DSCISEKLFMEMAELMVSEGWKDGAYEYLCIDDCWMAPOQDSEGRLOADPQRPFGIRQL 120							
Qy	121	ANYVHSKGLKGIYADYGNKTCAGFPGSFGYYDIDAQTFADWGVDDLKFDGICYCDSLENL 180							
Db	121	ANYVHSKGLKGIYADYGNKTCAGFPGSFGYYDIDAQTFADWGVDDLKFDGICYCDSLENL 180							
Qy	181	ADGYKHSLSALNRTGRSIVYSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240							
Db	181	ADGYKHSLSALNRTGRSIVYSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240							
Qy	241	SILDWTSFNQERIVDVAGPGWNPDPMLVIGNFGLSWNQVOTQALWAIMAAPLFMSNDL 300							
Db	241	SILDWTSFNQERIVDVAGPGWNPDPMLVIGNFGLSWNQVOTQALWAIMAAPLFMSNDL 300							
Qy	301	RHISPOAKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIIG 360							
Db	301	RHISPOAKALLQDKDVIATINQDPLGKQGYQLRQGDNFVWVERPLSGLAWAVAMINRQEIIG 360							
Qy	361	GPRSYYTIAVASLKGAVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417							
Db	361	GPRSYYTIAVASLKGAVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL 417							

Search completed: December 31, 2005, 23:48:33
Job time : 45.1379 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:16:51 ; Search time 33.5517 Seconds
(without alignments)
1195.837 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293

Sequence: 1 MQLRNPELHLCALALRFLA.....EWTSLRSHNPNTGTVLLQL 417

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2293	100.0	429	1 GBHUA	alpha-galactosidase
2	1806	78.8	419	2 JC4522	alpha-galactosidase
3	1051	45.8	405	2 S45522	alpha-N-acetylglucosaminidase
4	952.5	41.5	411	2 A35485	alpha-N-acetylglucosaminidase
5	930	40.6	358	2 A33265	alpha-N-acetylglucosaminidase
6	861	37.5	451	2 T24018	hypothetical protein
7	716.5	31.2	434	2 T47748	alpha-galactosidase
8	712	31.1	378	2 T50781	alpha-galactosidase
9	687.5	30.0	422	2 T06388	alpha-galactosidase
10	682	29.7	425	2 T10850	alpha-galactosidase
11	681.5	29.7	411	2 S07472	alpha-galactosidase
12	676	29.5	680	2 T36472	probable secreted
13	609	26.6	436	2 T39118	probable alpha-galactosidase
14	598	26.1	396	2 JC5558	alpha-galactosidase
15	557	24.3	469	2 S45453	alpha-galactosidase
16	543.5	23.7	545	2 S23582	alpha-galactosidase
17	521	22.7	471	2 S50312	alpha-galactosidase
18	519	22.6	471	2 S50311	alpha-galactosidase
19	516	22.5	444	2 S74221	alpha-galactosidase
20	514	22.4	471	2 T01021	alpha-galactosidase
21	510	22.2	471	2 S50310	alpha-galactosidase
22	502	21.9	471	1 GBBYAG	probable alpha-galactosidase
23	300.5	13.1	204	2 T04423	hypothetical protein
24	282.5	12.3	432	2 F83883	alpha-galactosidase
25	198	8.6	159	2 T04422	alpha-galactosidase
26	180.5	7.9	624	2 S74222	alpha-galactosidase
27	129	5.6	641	2 A55549	glucan 1,6-alpha-D-glucosyltransferase
28	125	5.5	348	2 A70311	hypothetical protein
29	116	5.1	4199	2 S76412	hypothetical protein

RESULT 1

GBHUA

alpha-galactosidase (EC 3.2.1.22) A precursor - human
N/Alternate names: alpha-D-galactoside galactohydrolase; melibiase
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1987 #sequence revision 27-Oct-1995 #text change 09-Jul-2004
C/Accession: S04081; A29608; A30214; S14879; A00896; B00896; I37140
R/Kornreich, R.; Desnick, R.J.; Bishop, D.F.

Nucleic Acids Res. 17, 3301-3302, 1989

A/Title: Nucleotide sequence of the human alpha-galactosidase A gene.

A/Reference number: S04081; MUID:89263745; PMID:2542896

A/Accession: S04081

A/Status: translation not shown

A/Molecule type: DNA

A/Residues: 1-429 <KOR>

A/Cross-references: UNIPROT:P06280; UNIPARC:UPI0000033A30; EMBL:X14448; NID:g31755; PIDN:R/Quinn, M.; Hantzopoulos, P.; Fidanza, V.; Calhoun, D.H.

Gene 58, 177-188, 1987

A/Title: A genomic clone containing the promoter for the gene encoding the human lysosom

A/Reference number: A29608; MUID:88112869; PMID:2892762

A/Accession: A29608

A/Molecule type: DNA

A/Residues: 1-64 <QUI>

A/Cross-references: UNIPARC:UPI000016A96A; GB:M18242; NID:g182944; PIDN:AAA52514.1; PID:R/Bishop, D.F.; Kornreich, R.; Desnick, R.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3903-3907, 1988

A/Title: Structural organization of the human alpha-galactosidase A gene: further eviden

A/Reference number: A30214; MUID:88234528; PMID:2836863

A/Accession: A30214

A/Molecule type: DNA

A/Residues: 1-64 <BIS>

A/Cross-references: UNIPARC:UPI000016A96A; EMBL:M20317; EMBL:J03249

R/Koide, T.; Ishiura, M.; Iwai, K.; Inoue, M.; Kaneda, Y.; Okada, Y.; Uchida, T.

FEBS Lett. 259, 353-356, 1990

A/Title: A case of Fabry's disease in a patient with no alpha-galactosidase A activity c

A/Reference number: S14879; MUID:90092580; PMID:2152885

A/Accession: S14879

A/Molecule type: mRNA

A/Residues: 1, K, 3-39, 'S', 41-429 <KOI>

A/Cross-references: UNIPARC:UPI000011E285; EMBL:X16889

A/Experimental source: Fabry's disease patient

R/Bishop, D.F.; Calhoun, D.H.; Bernstein, H.S.; Hantzopoulos, P.; Quinn, M.; Desnick, R.

Proc. Natl. Acad. Sci. U.S.A. 83, 4859-4863, 1986

A/Title: Human alpha-galactosidase A: nucleotide sequence of a cDNA clone encoding the m

A/Reference number: A00896; MUID:86259694; PMID:3014515

A/Accession: A00896

A/Molecule type: mRNA

A/Residues: 27-429 <BI2>

A/Cross-references: UNIPARC:UPI00000000358; GB:M13571; NID:g178245; PIDN:AAA51676.1; PID:R/Quinn, M.; Desnick, R.

A/Experimental source: lung

A/Accession: B00896

A/Molecule type: protein

A;Residues: 32-55,'S',57-58,'R',60-65,'S',67-68;228-232,'N',234-237,'A';298-326;'L',334-335;237-240;241-242;243-244;245-246;247-248;249-250;251-252;253-254;255-256;257-258;259-260;261-262;263-264;265-266;267-268;269-270;271-272;273-274;275-276;277-278;279-280;281-282;283-284;285-286;287-288;289-290;291-292;293-294;295-296;297-298;299-300;301-302;303-304;305-306;307-308;309-310;311-312;313-314;315-316;317-318;319-320;321-322;323-324;325-326;327-328;329-330;331-332;333-334;335-336;337-338;339-340;341-342;343-344;345-346;347-348;349-350;351-352;353-354;355-356;357-358;359-360;361-362;363-364;365-366;367-368;369-370;371-372;373-374;375-376;377-378;379-380;381-382;383-384;385-386;387-388;389-390;391-392;393-394;395-396;397-398;399-400;401-402;403-404;405-406;407-408;409-410;411-412;413-414;415-416;417-418;419-420;421-422;423-424;425-426;427-428;429-430;431-432;433-434;435-436;437-438;439-440;441-442;443-444;445-446;447-448;449-450;451-452;453-454;455-456;457-458;459-460;461-462;463-464;465-466;467-468;469-470;471-472;473-474;475-476;477-478;479-480;481-482;483-484;485-486;487-488;489-490;491-492;493-494;495-496;497-498;499-500;501-502;503-504;505-506;507-508;509-510;511-512;513-514;515-516;517-518;519-520;521-522;523-524;525-526;527-528;529-530;531-532;533-534;535-536;537-538;539-540;541-542;543-544;545-546;547-548;549-550;551-552;553-554;555-556;557-558;559-560;561-562;563-564;565-566;567-568;569-570;571-572;573-574;575-576;577-578;579-580;581-582;583-584;585-586;587-588;589-590;591-592;593-594;595-596;597-598;599-600;601-602;603-604;605-606;607-608;609-610;611-612;613-614;615-616;617-618;619-620;621-622;623-624;625-626;627-628;629-630;631-632;633-634;635-636;637-638;639-640;641-642;643-644;645-646;647-648;649-650;651-652;653-654;655-656;657-658;659-660;661-662;663-664;665-666;667-668;669-670;671-672;673-674;675-676;677-678;679-680;681-682;683-684;685-686;687-688;689-690;691-692;693-694;695-696;697-698;699-700;701-702;703-704;705-706;707-708;709-710;711-712;713-714;715-716;717-718;719-720;721-722;723-724;725-726;727-728;729-730;731-732;733-734;735-736;737-738;739-740;741-742;743-744;745-746;747-748;749-750;751-752;753-754;755-756;757-758;759-760;761-762;763-764;765-766;767-768;769-770;771-772;773-774;775-776;777-778;779-780;781-782;783-784;785-786;787-788;789-790;791-792;793-794;795-796;797-798;799-800;801-802;803-804;805-806;807-808;809-810;811-812;813-814;815-816;817-818;819-820;821-822;823-824;825-826;827-828;829-830;831-832;833-834;835-836;837-838;839-840;841-842;843-844;845-846;847-848;849-850;851-852;853-854;855-856;857-858;859-860;861-862;863-864;865-866;867-868;869-870;871-872;873-874;875-876;877-878;879-880;881-882;883-884;885-886;887-888;889-890;891-892;893-894;895-896;897-898;899-900;901-902;903-904;905-906;907-908;909-910;911-912;913-914;915-916;917-918;919-920;921-922;923-924;925-926;927-928;929-930;931-932;933-934;935-936;937-938;939-940;941-942;943-944;945-946;947-948;949-950;951-952;953-954;955-956;957-958;959-960;961-962;963-964;965-966;967-968;969-970;971-972;973-974;975-976;977-978;979-980;981-982;983-984;985-986;987-988;989-990;991-992;993-994;995-996;997-998;999-1000;1001-1002;1003-1004;1005-1006;1007-1008;1009-1010;1011-1012;1013-1014;1015-1016;1017-1018;1019-1020;1021-1022;1023-1024;1025-1026;1027-1028;1029-1030;1031-1032;1033-1034;1035-1036;1037-1038;1039-1040;1041-1042;1043-1044;1045-1046;1047-1048;1049-1050;1051-1052;1053-1054;1055-1056;1057-1058;1059-1060;1061-1062;1063-1064;1065-1066;1067-1068;1069-1070;1071-1072;1073-1074;1075-1076;1077-1078;1079-1080;1081-1082;1083-1084;1085-1086;1087-1088;1089-1090;1091-1092;1093-1094;1095-1096;1097-1098;1099-1100;1101-1102;1103-1104;1105-1106;1107-1108;1109-1110;1111-1112;1113-1114;1115-1116;1117-1118;1119-1120;1121-1122;1123-1124;1125-1126;1127-1128;1129-1130;1131-1132;1133-1134;1135-1136;1137-1138;1139-1140;1141-1142;1143-1144;1145-1146;1147-1148;1149-1150;1151-1152;1153-1154;1155-1156;1157-1158;1159-1160;1161-1162;1163-1164;1165-1166;1167-1168;1169-1170;1171-1172;1173-1174;1175-1176;1177-1178;1179-1180;1181-1182;1183-1184;1185-1186;1187-1188;1189-1190;1191-1192;1193-1194;1195-1196;1197-1198;1199-1200;1201-1202;1203-1204;1205-1206;1207-1208;1209-1210;1211-1212;1213-1214;1215-1216;1217-1218;1219-1220;1221-1222;1223-1224;1225-1226;1227-1228;1229-1230;1231-1232;1233-1234;1235-1236;1237-1238;1239-1240;1241-1242;1243-1244;1245-1246;1247-1248;1249-1250;1251-1252;1253-1254;1255-1256;1257-1258;1259-1260;1261-1262;1263-1264;1265-1266;1267-1268;1269-1270;1271-1272;1273-1274;1275-1276;1277-1278;1279-1280;1281-1282;1283-1284;1285-1286;1287-1288;1289-1290;1291-1292;1293-1294;1295-1296;1297-1298;1299-1300;1301-1302;1303-1304;1305-1306;1307-1308;1309-1310;1311-1312;1313-1314;1315-1316;1317-1318;1319-1320;1321-1322;1323-1324;1325-1326;1327-1328;1329-1330;1331-1332;1333-1334;1335-1336;1337-1338;1339-1340;1341-1342;1343-1344;1345-1346;1347-1348;1349-1350;1351-1352;1353-1354;1355-1356;1357-1358;1359-1360;1361-1362;1363-1364;1365-1366;1367-1368;1369-1370;1371-1372;1373-1374;1375-1376;1377-1378;1379-1380;1381-1382;1383-1384;1385-1386;1387-1388;1389-1390;1391-1392;1393-1394;1395-1396;1397-1398;1399-1400;1401-1402;1403-1404;1405-1406;1407-1408;1409-1410;1411-1412;1413-1414;1415-1416;1417-1418;1419-1420;1421-1422;1423-1424;1425-1426;1427-1428;1429-1430;1431-1432;1433-1434;1435-1436;1437-1438;1439-1440;1441-1442;1443-1444;1445-1446;1447-1448;1449-1450;1451-1452;1453-1454;1455-1456;1457-1458;1459-1460;1461-1462;1463-1464;1465-1466;1467-1468;1469-1470;1471-1472;1473-1474;1475-1476;1477-1478;1479-1480;1481-1482;1483-1484;1485-1486;1487-1488;1489-1490;1491-1492;1493-1494;1495-1496;1497-1498;1499-1500;1501-1502;1503-1504;1505-1506;1507-1508;1509-1510;1511-1512;1513-1514;1515-1516;1517-1518;1519-1520;1521-1522;1523-1524;1525-1526;1527-1528;1529-1530;1531-1532;1533-1534;1535-1536;1537-1538;1539-1540;1541-1542;1543-1544;1545-1546;1547-1548;1549-1550;1551-1552;1553-1554;1555-1556;1557-1558;1559-1560;1561-1562;1563-1564;1565-1566;1567-1568;1569-1570;1571-1572;1573-1574;1575-1576;1577-1578;1579-1580;1581-1582;1583-1584;1585-1586;1587-1588;1589-1590;1591-1592;1593-1594;1595-1596;1597-1598;1599-1600;1601-1602;1603-1604;1605-1606;1607-1608;1609-1610;1611-1612;1613-1614;1615-1616;1617-1618;1619-1620;1621-1622;1623-1624;1625-1626;1627-1628;1629-1630;1631-1632;1633-1634;1635-1636;1637-1638;1639-1640;1641-1642;1643-1644;1645-1646;1647-1648;1649-1650;1651-1652;1653-1654;1655-1656;1657-1658;1659-1660;1661-1662;1663-1664;1665-1666;1667-1668;1669-1670;1671-1672;1673-1674;1675-1676;1677-1678;1679-1680;1681-1682;1683-1684;1685-1686;1687-1688;1689-1690;1691-1692;1693-1694;1695-1696;1697-1698;1699-1700;1701-1702;1703-1704;1705-1706;1707-1708;1709-1710;1711-1712;1713-1714;1715-1716;1717-1718;1719-1720;1721-1722;1723-1724;1725-1726;1727-1728;1729-1730;1731-1732;1733-1734;1735-1736;1737-1738;1739-1740;1741-1742;1743-1744;1745-1746;1747-1748;1749-1750;1751-1752;1753-1754;1755-1756;1757-1758;1759-1760;1761-1762;1763-1764;1765-1766;1767-1768;1769-1770;1771-1772;1773-1774;1775-1776;1777-1778;1779-1780;1781-1782;1783-1784;1785-1786;1787-1788;1789-1790;1791-1792;1793-1794;1795-1796;1797-1798;1799-1800;1801-1802;1803-1804;1805-1806;1807-1808;1809-1810;1811-1812;1813-1814;1815-1816;1817-1818;1819-1820;1821-1822;1823-1824;1825-1826;1827-1828;1829-1830;1831-1832;1833-1834;1835-1836;1837-1838;1839-1840;1841-1842;1843-1844;1845-1846;1847-1848;1849-1850;1851-1852;1853-1854;1855-1856;1857-1858;1859-1860;1861-1862;1863-1864;1865-1866;1867-1868;1869-1870;1871-1872;1873-1874;1875-1876;1877-1878;1879-1880;1881-1882;1883-1884;1885-1886;1887-1888;1889-1890;1891-1892;1893-1894;1895-1896;1897-1898;1899-1900;1901-1902;1903-1904;1905-1906;1907-1908;1909-1910;1911-1912;1913-1914;1915-1916;1917-1918;1919-1920;1921-1922;1923-1924;1925-1926;1927-1928;1929-1930;1931-1932;1933-1934;1935-1936;1937-1938;1939-1940;1941-1942;1943-1944;1945-1946;1947-1948;1949-1950;1951-1952;1953-1954;1955-1956;1957-1958;1959-1960;1961-1962;1963-1964;1965-1966;1967-1968;1969-1970;1971-1972;1973-1974;1975-1976;1977-1978;1979-1980;1981-1982;1983-1984;1985-1986;1987-1988;1989-1990;1991-1992;1993-1994;1995-1996;1997-1998;1999-2000;2001-2002;2003-2004;2005-2006;2007-2008;2009-2010;2011-2012;2013-2014;2015-2016;2017-2018;2019-2020;2021-2022;2023-2024;2025-2026;2027-2028;2029-2030;2031-2032;2033-2034;2035-2036;2037-2038;2039-2040;2041-2042;2043-2044;2045-2046;2047-2048;2049-2050;2051-2052;2053-2054;2055-2056;2057-2058;2059-2060;2061-2062;2063-2064;2065-2066;2067-2068;2069-2070;2071-2072;2073-2074;2075-2076;2077-2078;2079-2080;2081-2082;2083-2084;2085-2086;2087-2088;2089-2090;2091-2092;2093-2094;2095-2096;2097-2098;2099-2100;2101-2102;2103-2104;2105-2106;2107-2108;2109-2110;2111-2112;2113-2114;2115-2116;2117-2118;2119-2120;2121-2122;2123-2124;2125-2126;2127-2128;2129-2130;2131-2132;2133-2134;2135-2136;2137-2138;2139-2140;2141-2142;2143-2144;2145-2146;2147-2148;2149-2150;2151-2152;2153-2154;2155-2156;2157-2158;2159-2160;2161-2162;2163-2164;2165-2166;2167-2168;2169-2170;2171-2172;2173-2174;2175-2176;2177-2178;2179-2180;2181-2182;2183-2184;2185-2186;2187-2188;2189-2190;2191-2192;2193-2194;2195-2196;2197-2198;2199-2200;2201-2202;2203-2204;2205-2206;2207-2208;2209-2210;2211-2212;2213-2214;2215-2216;2217-2218;2219-2220;2221-2222;2223-2224;2225-2226;2227-2228;2229-2230;2231-2232;2233-2234;2235-2236;2237-2238;2239-2240;2241-2242;2243-2244;2245-2246;2247-2248;2249-2250;2251-2252;2253-2254;2255-2256;2257-2258;2259-2260;2261-2262;2263-2264;2265-2266;2267-2268;2269-2270;2271-2272;2273-2274;2275-2276;2277-2278;2279-2280;2281-2282;2283-2284;2285-2286;2287-2288;2289-2290;2291-2292;2293-2294;2295-2296;2297-2298;2299-2300;2301-2302;2303-2304;2305-2306;2307-2308;2309-2310;2311-2312;2313-2314;2315-2316;2317-2318;2319-2320;2321-2322;2323-2324;2325-2326;2327-2328;2329-2330;2331-2332;2333-2334;2335-2336;2337-2338;2339-2340;2341-2342;2343-2344;2345-2346;2347-2348;2349-2350;2351-2352;2353-2354;2355-2356;2357-2358;2359-2360;2361-2362;2363-2364;2365-2366;2367-2368;2369-2370;2371-2372;2373-2374;2375-2376;2377-2378;2379-2380;2381-2382;2383-2384;2385-2386;2387-2388;2389-2390;2391-2392;2393-2394;2395-2396;2397-2398;2399-2400;2401-2402;2403-2404;2405-2406;2407-2408;2409-2410;2411-2412;2413-2414;2415-2416;2417-2418;2419-2420;2421-2422;2423-2424;2425-2426;2427-2428;2429-2430;2431-2432;2433-2434;2435-2436;2437-2438;2439-2440;2441-2442;2443-2444;2445-2446;2447-2448;2449-2450;2451-2452;2453-2454;2455-2456;2457-2458;2459-2460;2461-2462;2463-2464;2465-2466;2467-2468;2469-2470;2471-2472;2473-2474;2475-2476;2477-2478;2479-2480;2481-2482;2483-2484;2485-2486;2487-2488;2489-2490;2491-2492;2493-2494;2495-2496;2497-2498;2499-2500;2501-2502;2503-2504;2505-2506;2507-2508;2509-2510;2511-2512;2513-2514;2515-2516;2517-2518;2519-2520;2521-2522;2523-2524;2525-2526;2527-2528;2529-2530;2531-2532;2533-2534;2535-2536;2537-2538;2539-2540;2541-2542;2543-2544;2545-2546;2547-2548;2549-2550;2551-2552;2553-2554;2555-2556;2557-2558;2559-2560;2561-2562;2563-2564;2565-2566;2567-2568;2569-2570;2571-2572;2573-2574;2575-2576;2577-2578;2579-2580;2581-2582;2583-2584;2585-2586;2587-2588;2589-2590;2591-2592;2593-2594;2595-2596;2597-2598;2599-2600;2601-2602;2603-2604;2605-2606;2607-2608;2609-2610;2611-2612;2613-2614;2615-2616;2617-2618;2619-2620;2621-2622;2623-2624;2625-2626;2627-2628;2629-2630;2631-2632;2633-2634;2635-2636;2637-2638;2639-2640;2641-2642;2643-2644;2645-2646;2647-2648;2649-2650;2651-2652;2653-2654;2655-2656;2657-2658;2659-2660;2661-2662;2663-2664;2665-2666;2667-2668;2669-2670;2671-2672;2673-2674;2675-2676;2677-2678;2679-2680;2681-2682;2683-2684;2685-2686;2687-2688;2689-2690;2691-2692;2693-2694;2695-2696;2697-2698;2699-2700;2701-2702;2703-2704;2705-2706;2707-2708;2709-2710;2711-2712;2713-2714;2715-2716;2717-2718;2719-2720;2721-2722;2723-2724;2725-2726;2727-2728;2729-2730;2731-2732;2733-2734;2735-2736;2737-2738;2739-2740;2741-2742;2743-2744;2745-2746;2747-2748;2749-2750;2751-2752;2753-2754;2755-2756;2757-2758;2759-2760;2761-2762;2763-2764;2765-2766;2767-2768;2769-2770;2771-2772;2773-2774;2775-2776;2777-2778;2779-2780;2781-2782;2783-2784;2785-2786;2787-2788;2789-2790;2791-2792;2793-2794;2795-2796;2797-2798;2799-2800;2801-2802;2803-2804;2805-2806;2807-2808;2809-2810;2811-2812;2813-2814;2815-2816;2817-2818;2819-2820;2821-2822

C;Genetics:
A;Gene: GDB:NAGA
A;Cross-references: GDB:I19445; OMIM:104170
A;Map position: 22q11-22q11
A;Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3; 367/3
C;Superfamily: alpha-galactosidase
C;Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F;1-17/Domain: signal sequence #status predicted <Sig>
F;18-41/Product: alpha-N-acetylgalactosaminidase, splice form 2 #status predicted <MAT>
F;124,177,201,359,385,391/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 41.5%; Score 952.5; DB 2; Length 411;
Best Local Similarity 50.0%; Pred. No. 6.7e-74;
Matches 198; Conservative 55; Mismatches 120; Indels 23; Gaps 9;

QY 32 LDNGLTPTTGWTLHWFRCMNCIDCBEPDSCISSEKLFMEAEALMVSEGWKADAGEYLICI 91
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 18 LDNGLLQTTPPGWLAWERFCNICINCEDEPKICISEQLFMEMADRMAQGWRDMGYTYLNI 77
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 92 DDCWMAFORSEGRLOAQDPFPFHGIROLANTVHSVKGKLKIYADVGNKTCAFGPPG-SFG 150
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 78 DDWCIG-GRDASGRIMPDKPFFPHGIPFLADVVHSLGLKIYADMGNFTCMGYPGTTLD 136
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 151 YYDIDAOTCFADMGVDLLKFDCYCDSLENLAGDYKHMSLANRTGRSIVVYSCWELPMMP 210
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 137 KVVODAQTFAEVKVMDLKDGCF-STPERAQGYFKMAAALNATGPFAFCSWPAYEGG 195
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 211 F-QKPNTTEIRQYNHNRFNFADIDDWSKSILDWTSFNQBRIVDVAGPGWNDDPMLV 269
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 196 LPRVNYSLLADICNLNRNYDDIQDSWHSVLINFWFEHQDILOPVAGFGHWNDPDL 255
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 270 IGNFLGSNNQQVTQVALWAIPAFLMSNDLRHISPQAKALLQDKDVIAINQDPLGKOGY 329
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 256 IGFGLSIEDSRQAQWALTVAAPLMSTMSTDLRTISAQNMDILQNPLMIKINQDPLGIQGR 315
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 330 QL-RQDNFWEWERPLSGLA WAMIN-ROETGGRPSSYTIAVASLGKVACNPA CFITOL 387
 :::: |:::: |:::: |:::: |:::: |:::: |:::: |:::: |:::: |
Db 316 RIHKESLI EYMTPLSNKASLVFFSCRTDM--PYRHSSLGQLN-----FTGSV 364
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

QY 388 LPVKRKLGFYEWTSLRLSH-----INPTGTVLLOL 417
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Db 365 IYEADTVSGDIIISGLRDTHFTVIINSGVVMWYL 400
 ||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :

RESULT 5
A33265
alpha-N-acetylgalactosaminidase (EC 3.2.1.49) precursor, splice form 1 - human
N;Alternate names: alpha-galactosidase B
C;Species: Homo sapiens (man)
C;Date: 21-Feb-1990 #sequence_revision 21-Feb-1990 #text_change 09-Jul-2004
C;Accession: A33265
R;Tsuiji, S.; Yamauchi, T.; Hiraiwa, M.; Isobe, T.; Okuyama, T.; Sakimura, K.; Takahashi,
Biochem. Biophys. Res. Commun. 163, 1498-1504, 1989
A;Title: Molecular cloning of a full-length cDNA for human alpha-N-acetylgalactosaminida
A;Reference number: A33265; MUID:89392067; PMID:2551294
A;Accession: A33265
A;Molecule type: mRNA
A;Residues: 1-358 <TSU>
A;Cross-references: UNIPROT:P17050; UNIPARC:UPI000016ADB4; GB:M29276; NID:g189052; PIDN:
A;Experimental source: clone pcD-HS1204
C;Genetics:
A;Gene: GDB:NAGA
A;Cross-references: GDB:I19445; OMIM:104170
A;Map position: 22q11-22q11
A;Introns: 6/1; 51/2; 108/3; 168/1; 199/3; 253/3; 319/3; 367/3
C;Superfamily: alpha-galactosidase
C;Keywords: alternative splicing; glycoprotein; glycosidase; hydrolase; lysosome
F;1-17/Domain: signal sequence #status predicted <Sig>
F;18-358/Product: alpha-N-acetylgalactosaminidase, splice form 1 #status predicted <MAT>
F;124,177,201/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 40.6%; Score 930; DB 2; Length 358;
Best Local Similarity 57.7%; Pred. No. 4.8e-72;

A:Accession: T50781
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-378 <ZHU>
A:Cross-references: UNIPROT:Q42656; UNIPARC:UPI00001256A8; EMBL:L27992; PID:AAA33022.1
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 31.1%; Score 712; DB 2; Length 378;
Best Local Similarity 42.5%; Pred. No. 2.7e-53;
Matches 152; Conservative 51; Mismatches 113; Indels 42; Gaps 8;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEFDCISEKLFMEAEIMVSEGWKDGAGYEYLCI 91
DB 16 LANGLGLTPPMGWSNHFRCNLD-----EKLLRETADAMVSKGLAALGYKYINL 65

QY 92 DDCWMAPODSEGRLOADQPPHGIROLANYVHSGKLGKGIYADVGNKTCAGPFGSFG 150
DB 66 DDCWALNRDSQGNLVKPSGIFPSGIGKALADYVHSGKLGKGIYDAGTGTCTSKTWPGLG 125

QY 151 YYDIDAQTFADWGVDLLKFDGCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 126 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSKALLNSGRSIFFSICEWG---- 180

QY 210 PFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFNOERIVDVAGPGGWNNDP 265
DB 181 --EEDPATWAEVGVNSWRTTGDIDDSWSMTSRADMDNDKWASY-----AGPGGWNNDP 230

QY 266 DMLVGNFGLSNQOVTQMAWAIWAAPLFWNSDLRHISPOKALLQDKDVIAINODPLG 325
DB 231 DMLVNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRSDMGATFQLLSNAEVIANNQDKLG 290

QY 326 KQGYQLRQGDNFVEWPERPLSGLAWAVAMINR-----QEIGGPRSYTIAVASL 372
DB 291 VQGNKVYTGDLVWAGPLSGKRAVALWNRGSSATITAYWSDVGLPSTAVVNRDL 348

RESULT 9
T06388
alpha-galactosidase (EC 3.2.1.22) - soybean
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
C:Accession: T06388
R:Davis, M.O.; Walker, J.C.; Smith, D.
A:Description: Cloning and expression of a soybean alpha galactosidase gene.
A:Reference number: Z15645
A:Accession: T06388
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-422 <DAV>
A:Cross-references: UNIPROT:Q39811; UNIPARC:UPI00000A7200; EMBL:U12926; PID:g927574; PID:g927574;
A:Experimental source: strain williams
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 30.0%; Score 687.5; DB 2; Length 422;
Best Local Similarity 43.2%; Pred. No. 4e-51;
Matches 143; Conservative 45; Mismatches 112; Indels 31; Gaps 7;

QY 32 LDNGLARTPTMGWLHWFRCNLDQCBEFDCISEKLFMEAEIMVSEGWKDGAGYEYLCI 91
DB 60 LDNGLGHPTPMGWSNHFACN-----IKEDLLRETADAMVSTGLAALGYQYINI 109

QY 92 DDCWMAPODSEGRLOADQPPHGIROLANYVHSGKLGKGIYADVGNKTCAGPFGSFG 150
DB 110 DDCWELNRDSKGNLVKPASTPPSGMKALADYVHSGKLGKGIYDAGTGTCTSKTWPGLG 169

QY 151 YYDIDAQTFADWGVDLLKFDGCYCDLSLENLADGYKHMSLALNRTGRSIVYS-CEWPLYMW 209
DB 170 HEEQDAKTFASGVVDYLYKDYCNNNNI-SPKERYPIMSEALANTGRPIFFSLCEWG---- 224

QY 210 PFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFNOERIVDVAGPGGWNNDP 265
DB 225 --SEDPATWAKSVGNSWRTTGTIDQKWDMSIRADLNDKWASY-----AGPGGWNNDP 274

QY 266 DMLVGNFGLSNQOVTQMAWAIWAAPLFWNSDLRHISPOKALLQDKDVIAINODPLG 325
DB 275 DMLVNGGMMTTEYRSHFSIWAALAKAPLLIGCDIRALDATTKELLSNKEVIANNQDKLG 334

QY 326 KQGYQLRQGDNFVEWPERPLSGLAWAVAMINR 356
DB 335 VQGNKVSTNDLEWVAGPLSGNKKVAVILWNR 365

RESULT 10
T10860
alpha-galactosidase (EC 3.2.1.22) - kidney bean
C:Species: Phaseolus vulgaris (kidney bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T10860
R:Davis, M.O.; Walker, J.C.; Smith, D.
A:Description: Cloning and expression of a pinto bean alpha galactosidase gene.
A:Reference number: Z17189
A:Accession: T10860
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-425 <DAV>
A:Cross-references: UNIPROT:Q41100; UNIPARC:UPI00000A0F13; EMBL:U12927; PID:g927576; PID:g927576;
C:Function:
A:Description: catalyzes hydrolysis of melibiose into galactose and glucose
C:Superfamily: alpha-galactosidase
C:Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 682; DB 2; Length 425;
Best Local Similarity 41.5%; Pred. No. 1.2e-50;
Matches 144; Conservative 54; Mismatches 113; Indels 36; Gaps 9;

QY 19 LALVSWDIPGARAL-DNGLARTPTMGWLHWFRCNLDQCBEFDCISEKLFMEAEIMV 77
DB 49 MMMHSREVDHRRNLVGNGLGQTPPMGWSNHFACN-----INEDLIRETADAMV 98

QY 78 SEGKADAGYELCIDDCWMAPODSEGRLOADQPPHGIROLANYVHSGKLGKGIYADV 137
DB 99 STGLAALGYQYINIDDCWELNRDSQGNLVKPASTPPSGMKALADYVHSGKLGKGIYSDA 158

QY 138 GNKTCAGPFGSFGYDIDAQTFADWGVDLLKFDGCYCDLSLENLA--DGYKHMSLALNRT 194
DB 159 GTQCSKTMPSGLGHEBQDAKTFASWIDYLYKDYCN-----ENKNISPKERYPPMSKALANS 215

QY 195 GRSIVYS-CEWPLYMWPFQKPNYTEIRQYCNHWRNFADIDDSWKSISKILD-----WTSFN 249
DB 216 GRPIFFSLCEWG-----SEDPATWAKSVGNSWRTTGTIDKWMESIRADLNDSEWASY- 268

QY 250 QERIVDVAGPGGWNNDPMLVGNFGLSNQOVTQMAWAIWAAPLFWNSDLRHISPOKA 309
DB 269 -----AGPGGWNNDPMLVGNFGLSNQOVTTEYRSHFSIWAALAKAPLLIGCDIRALDVTTCB 321

QY 310 LLQDKDVIAINODPLGKQGYQLRQGDNFVEWPERPLSGLAWAVAMINR 356
DB 322 LLSNEEVIANNQDKLGVGKVKNSNNDLEWVAGPLSGNRLAVILWNR 368

RESULT 11
S07472
alpha-galactosidase (EC 3.2.1.22) precursor - guar
C:Species: Cyamopsis tetragonoloba (guar, cluster bean)
C:Date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 09-Jul-2004
C:Accession: S07472
R:Overbeek, N.; Fellinger, A.J.; Toonen, M.Y.; van Wassenaar, D.; Verrips, C.T.
Plant Mol. Biol. 13, 541-550, 1989
A:Title: Cloning and nucleotide sequence of the alpha-galactosidase cDNA from Cyamopsis
A:Reference number: S07472; MUID:91370836; PMID:2577496

A;Accession: S07472
A;Molecule type: mRNA
A;Residues: 1-411 <OV>
A;Cross-references: UNIPROT:P14749; UNIPARC:UPI00001256A9; EMBL:X14619; NID:g18291; PIDN
A;Note: the authors translated the codon ATT for residue 20 as Asn, TAT for residue 140
C;Superfamily: alpha-galactosidase
C;Keywords: glycosidase; hydrolase

Query Match 29.7%; Score 681.5; DB 2; Length 411;
Best Local Similarity 42.2%; Pred. No. 1.3e-50;
Matches 141; Conservative 52; Mismatches 102; Indels 39; Gaps 9;
QY 33 DNGIARTTGMGLHWRFCMLDQCQEPDSCISEKLFMEAEIWMVSEGWKAGYEYLCID 92
DB 49 ENGLGQTTPMGWNSNNHFGCD-----INENVRETADAMVSTGLAALGQYINLD 98
QY 93 DCWMAPOQDSRGRIQADPQRPFGHGIROLANYVHSGKLKLGIVADYGVNKTCA-GPFGSGFY 151
DB 99 DCWAEIARNDSEGNMVPNAAPFSGIKALADYVHSGKLKLGIVSDAGNQTCRKMPGSLGH 158
QY 152 YDIDAQTFADWGLLKFDGVCYCDLENLA---DGYKHMSLALNRTGRSIVYS-CEWPL 206
DB 159 EEQAKTTPASGVVLYKYDNC-----ENLGISVKEVPPMGKALLSSGRPIFFSWCE--- 210
QY 207 YMWPFQKPNYTEIROQYCNHWNFNADIDDSWKSIIKSLD---WTSFNOERIVDVAGPGGW 262
DB 211 --WGVEDPQIW-AKSIGNSWTTGDIEDNWNMTSIADSNKMASY-----AGPGGW 259
QY 263 NDPMLVIGNFGLSNQVQTMALWAIWAAPLFMSNDLRHISPOAKALLQDKQVIAINQD 322
DB 260 NDPMLVGVNGMWTTEYRSHFSIWALAKAPLLVGCIDIRAMDDTTHELISNAEVIANNQD 319
QY 323 PLGKQGYQLROGDNFEVWERPLSGLAWAVAMINR 356
DB 320 KLGVGQKKVYSTNDLEVWAGPLSONKVAIVLWNR 353

RESULT 12

T36472
probable secreted alpha-galactosidase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T36472
R;Seeger, K.J.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21607
A;Accession: T36472
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-680 <SEE>
A;Cross-references: UNIPROT:Q9S2C9; UNIPARC:UPI00000DB403; EMBL:AL110470; PIDN:CAB54169
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCODB:SCF95.12

Query Match 29.5%; Score 676; DB 2; Length 680;
Best Local Similarity 38.2%; Pred. No. 7.3e-50;
Matches 161; Conservative 67; Mismatches 140; Indels 54; Gaps 14;
QY 11 GCALALRFALVSWDIPGARALDNGIARTPTMGWLHWRFCMLDQCQEPDSCISEKLFM 70
DB 29 GLAVSAQVAPAAAPDPAPASSGHEQLALTTPMGFNWNN---STHCRDE---FNESMVK 80
QY 71 EMAEIWMSEGWKAGYEYLCIDDCWMAPOQDSRGRIQADPQRPFGHGIROLANYVHSGKLK 130
DB 81 GIADLFVERGLKAGYEYVNLDDCWALPERDADGRLVDPKRPFGNGIEAVADYVHSGKLK 140
QY 131 LGIYADVGNKTCA--GPFGSGFYVDIDAQTFADWGLLKFDGVCYCDLENLADGYKMS 188
DB 141 FGIYTSAGTKTCSSIGFPEALGHEYSDAQFADWGVLYKYDNCNNQGV--AKORYTMR 199
QY 189 LALNRTGRSIVYS-CEW---PLYMWPQKPNYTEIROQYCNHWNFNADIDDSWKSIIKSL 243

DB 200 DALAATGRPIVYSICEWGENKP---WEWA---GDLQQL---WRTTGDINDSWGSMSSIM 249
QY 244 DWTSTFQNERIVDVAGPGGWNDDPMLVIGNFGLSNQVQTMALWAIWAAPLFMSNDLRHI 303
DB 250 K-----SNLEAEYARPGWNDPDMLEVGNGMTTEYRTHFSWMSIWAAPLLIGTDLRATA 305
QY 304 SPOAKALLQDKQVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINROEIOGPR 363
DB 306 PESAFEILTNDVEIVADQDPLGKQGEVVSSEGGWVVSKEIADGSGRAVALFNE-----GSR 361
QY 364 SYTTAVASLKGVCACNPACFITQLLPVKRLKGFYE-W-----TSRLRSHINTGTGVLL 415
DB 362 AORIETTAQAVG-----LPKSRGYTMRDLWKHSDTNTTGTGIAATVPAHGTVLV 409
QY 416 QL 417
DB 410 RV 411

RESULT 13

T39118
probable alpha-galactosidase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T39118
R;Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A;Reference number: Z21829
A;Accession: T39118
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-436 <HUN>
A;Cross-references: UNIPROT:Q9URZ0; UNIPARC:UPI000000BE8C; EMBL:AL132779; PIDN:CAB60017
A;Experimental source: strain 972h-; cosmid c869
C;Genetics:
A;Gene: SPDB:SPAC869.07c
A;Map position: 1
C;Superfamily: alpha-galactosidase

Query Match 26.6%; Score 609; DB 2; Length 436;
Best Local Similarity 35.8%; Pred. No. 2.3e-44;
Matches 136; Conservative 65; Mismatches 119; Indels 60; Gaps 14;
QY 10 LGCALALRFALVSWDIPGARALDNGIARTPTMGWLHWRFCMLDQCQEPDSCISEKLF 69
DB 7 LNC-FFLVFLFLPSPDVHGSY---NGLGLKPEQMGWNSWKYACDID-----ESII 52
QY 70 MEMAEIWMSEGWKAGYEYLCIDDCWMAPOQDSRGRIQADPQRPFGHGIROLANYVHSGK 128
DB 53 LNNAKAIKEEGLLDLGYEYIVMDDCWKSHERNATTTGRLEANPDKFPNGIGSMAXKGLHDMG 112
QY 129 LKGIYADVGNKTCA--GPFGSGFYVDIDAQTFADWGLLKFDGVCYCDLENLA---DGY 184
DB 113 FKGMYSAGKYTCAGFPGLNHEQIDADTFADWGVLYKYDNCFNKSGKGVPLISVRY 172
QY 185 KHMSLALNRTGRSIVYS-CEW-PLYMWPQKPNYTEIROQYCNHWNFNADIDDSWK----- 237
DB 173 KEMSDALNKTGRPIFYSLCQWGEDFVNWG-----NTIANSWRISGDIFFTFSRKDVR 225
QY 238 -----SIKSILDWTSFNQERIVDVAG-PGWNDDPDMVLVGNFGLSNNQ 280
DB 226 CPCETIECFALQGDHCSVMNIIISKASFLSSK---AGMNSGWNDLDSLEVNGSGMSFEY 281
QY 281 VTQWALWAIWAAPLFMSNDLRHISPOAKALLQDKQVIAINQDPLGKQ-----YOLRQDN 336
DB 282 KTHFTWAILKSPILINDVSSMSPMOKLIIVSNKELISINQD-IGTNPAALINWKKYGYDE 340
QY 337 F-EVWERPLSGLAWAVAMIN 355
DB 341 YIELFSGELSNNDWVAVLN 360

RESULT 14


```

C:Superfamily: alpha-galactosidase
C:Keywords: glycoprotein; glycosidase; hydrolase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-469/Product: alpha-galactosidase MEL #status predicted <MAT>

Query Match          24.3%; Score 557; DB 2; Length 469;
Best Local Similarity 35.3%; Pred. No. 7.4e+00;
Matches 134; Conservative 52; Mismatches 128; Indels 66; Gaps 12;

QY 18 FLALV--SWDIFGARALDNLGARTPTMGWLHREFMNCNLDCQEPDSCISEKLFMEMAEL 75
DB 5 FYALFFSSTDVLAASPSYNGLGITPOMGNDNWNNSFGCS-----VKBELLLGTAEK 54

QY 76 MYSEGWQAGAYEYLCIDCCWMAPOQRDEGRLOADPQRFPHGIRQLANYVHSGKLGLGIYA 135
DB 55 IVKLGKDLGYNIIILDDCW--SSGRSSNSGLLADDSKFFPHGMKYVAEQLHNSQLKFGMYS 113

QY 136 DVGNKTCAGFPGSGFYIIDATQTFADMGVDLLKFGDCY----CDSLNIENADGKYKHSAL 191
DB 114 SAGEVTCAGYAGSLGYEDMDAATFASWDVLYLKDNCYNKGFETPEISYKRYKAMSDAL 173

QY 192 NRTGRSIYVS--CEWP---LYMHPFQKPNYTEIROYCNHWRNFADI-----232
DB 174 NKTGRPIFYSLCNWGQDLTFYWG-----SAISNSWRMSGDVYPOFDRPDSRCPCSG 224

QY 233 ----DSRWK----SIKSLDWTFSNOERIYDVAGPGGWNPDMLVIGNFGLSNQQVTQMA 285
DB 225 DRYDCSYPGFHCYSIWNILNKAAPMGQN-----AAPGGNLDMLLEVGVGNMSDSEEVAHFS 280

QY 286 LWAIMAAPLFMSNDLHRHSIPQAKALLQDXVTAINDPLGKQG-----YQLRGQDNF--- 337
DB 281 MWAIVKSPLIIGADIDDLKDSLSVYSNPAVTAINDQVLGTPATRIWKYHVSDDQYGE 340

QY 338 --EVMERPSGLAWAVAMIN 355
DB 341 ETQLWSGPLDNGHVVALLN 360

Search completed: December 31, 2005, 23:47:42
Job time : 36.5517 secs

```

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2006 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 31, 2005, 23:47:17 ; Search time 122.703 Seconds
(without alignments)
1419.967 Million cell updates/sec

Title: US-10-602-220-16

Perfect score: 2293
Sequence: 1 MQLRNPETHGICALALRFLA.....EWTSLRSHINPTGTVLLQL 417

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 s

```
Database : Published Applications_AA_Main:*
1: /csgn2_6/prodata1/pubpaa/US07_PUBCOMB.psp:*
2: /csgn2_6/prodata1/pubpaa/US08_PUBCOMB.psp:*
3: /csgn2_6/prodata1/pubpaa/US09_PUBCOMB.psp:*
4: /csgn2_6/prodata1/pubpaa/US10A_PUBCOMB.psp:*
5: /csgn2_6/prodata1/pubpaa/US10B_PUBCOMB.psp:*
6: /csgn2_6/prodata1/pubpaa/US11_PUBCOMB.psp:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		Length	DB	ID	Description
		Query Match					
1	2293	100.0	417	3	US-09-993-059-16		Sequence 16, Appl
2	2293	100.0	417	4	US-10-103-327-16		Sequence 16, Appl
3	2293	100.0	417	4	US-10-602-219-16		Sequence 16, Appl
4	2293	100.0	417	4	US-10-602-220-16		Sequence 16, Appl
5	2293	100.0	417	5	US-10-851-388-16		Sequence 16, Appl
6	2293	100.0	417	5	US-10-984-389-16		Sequence 16, Appl
7	2293	100.0	421	3	US-09-993-059-12		Sequence 12, Appl
8	2293	100.0	421	4	US-10-103-327-12		Sequence 12, Appl
9	2293	100.0	421	4	US-10-602-219-12		Sequence 12, Appl
10	2293	100.0	421	4	US-10-602-220-12		Sequence 12, Appl
11	2293	100.0	421	5	US-10-851-388-12		Sequence 12, Appl
12	2293	100.0	421	5	US-10-984-389-12		Sequence 12, Appl
13	2293	100.0	423	3	US-09-993-059-18		Sequence 18, Appl
14	2293	100.0	423	4	US-10-103-327-18		Sequence 18, Appl
15	2293	100.0	423	4	US-10-602-219-18		Sequence 18, Appl
16	2293	100.0	423	4	US-10-602-220-18		Sequence 18, Appl
17	2293	100.0	423	5	US-10-851-388-18		Sequence 18, Appl
18	2293	100.0	423	5	US-10-984-389-18		Sequence 18, Appl
19	2293	100.0	427	3	US-09-993-059-14		Sequence 14, Appl
20	2293	100.0	427	4	US-10-103-327-14		Sequence 14, Appl
21	2293	100.0	427	4	US-10-602-219-14		Sequence 14, Appl
22	2293	100.0	427	4	US-10-602-220-14		Sequence 14, Appl
23	2293	100.0	427	5	US-10-851-388-14		Sequence 14, Appl
24	2293	100.0	427	5	US-10-984-389-14		Sequence 14, Appl
25	2293	100.0	429	4	US-10-411-037-68		Sequence 68, Appl
26	2293	100.0	429	4	US-10-411-026-68		Sequence 68, Appl
27	2293	100.0	429	4	US-10-410-962-68		Sequence 68, Appl

28	2293	100.0	429	4	US-10-411-049-68	Sequence 68, Appl
29	2293	100.0	429	4	US-10-410-930-68	Sequence 68, Appl
30	2293	100.0	429	4	US-10-410-937-68	Sequence 68, Appl
31	2293	100.0	429	4	US-10-411-012-68	Sequence 68, Appl
32	2293	100.0	429	4	US-10-411-913-68	Sequence 68, Appl
33	2293	100.0	429	5	US-10-370-715B-278	Sequence 278, Appl
34	2293	100.0	429	5	US-10-410-980-68	Sequence 68, Appl
35	2293	100.0	429	5	US-10-410-897-68	Sequence 68, Appl
36	2293	100.0	431	3	US-09-993-059-10	Sequence 10, Appl
37	2293	100.0	431	4	US-10-103-327-10	Sequence 10, Appl
38	2293	100.0	431	4	US-10-602-219-10	Sequence 10, Appl
39	2293	100.0	431	4	US-10-602-226-10	Sequence 10, Appl
40	2293	100.0	431	5	US-10-851-388-10	Sequence 10, Appl
41	2293	100.0	431	5	US-10-984-389-10	Sequence 10, Appl
42	2293	100.0	435	3	US-09-993-059-6	Sequence 6, Appl
43	2293	100.0	435	4	US-10-103-327-6	Sequence 6, Appl
44	2293	100.0	435	4	US-10-602-219-6	Sequence 6, Appl
45	2293	100.0	435	4	US-10-602-220-6	Sequence 6, Appl

ALIGNMENTS

```

RESULT 1
US-09-993-059-16
? Sequence 16, Application US/09993059
? Publication No. US20020088024A1
? GENERAL INFORMATION:
? APPLICANT: GARGER, Stephen A.
? APPLICANT: TURPEN, Thomas H.
? APPLICANT: KUMAGAI, Monto H.
? TITLE OF INVENTION: PROTO H.
? TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
? FILE REFERENCE: 008010087CEUS06
? CURRENT APPLICATION NUMBER: US/09/993,059
? CURRENT FILING DATE: 2001-11-13
? NUMBER OF SEQ ID NOS: 37
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 16
? LENGTH: 417
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-993-059-16

```

[illegible]

Db 361 GPRSYYTIAVASLKGKGVACNACFITTQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 2

US-10-103-327-16
; Sequence 16, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-16

Query Match 100.0%; Score 2293; DB 4; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MQLRNPGLHGCALALRFLALVSDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60
Db	1	MQLRNPGLHGCALALRFLALVSDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60
Qy	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL	120
Db	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL	120
Qy	121	ANYVHSGKLGKIYADVGNKTCAGFGSGFYIIDDAQTADGWGVDLLKFDGCGYCDLSLENL	180
Db	121	ANYVHSGKLGKIYADVGNKTCAGFGSGFYIIDDAQTADGWGVDLLKFDGCGYCDLSLENL	180
Qy	181	ADGYKMSLALNTRGSRIVSYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240
Db	181	ADGYKMSLALNTRGSRIVSYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240
Qy	241	SILDWTSFNQERIVDVAGPGGWNDDPMLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300
Db	241	SILDWTSFNQERIVDVAGPGGWNDDPMLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300
Qy	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSLGLAWAVAMINRQEI	360
Db	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSLGLAWAVAMINRQEI	360
Qy	361	GPRSYYTIAVASLKGKGVACNACFITTQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417
Db	361	GPRSYYTIAVASLKGKGVACNACFITTQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417

RESULT 3

US-10-602-219-16
; Sequence 16, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: POGUE, Gregory P.
; APPLICANT: ERWIN, Robert L.
; APPLICANT: GRILL, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYSSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219

; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-16

Query Match	100.0%;	Score 2293;	DB 4;	Length 417;
Best Local Similarity	100.0%;	Pred. No. 9e-220;		
Matches	417;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MQLRNPGLHGCALALRFLALVSDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60	
Db	1	MQLRNPGLHGCALALRFLALVSDIPGARALDGLARTPTMGWLHWRFCMNLDCOEEP	60	
Qy	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL	120	
Db	61	DSICSEKLFMEAEMLVSEGWKADAGYEYLCIDCWMAPOQDSGRLOADPQRPFGIRQL	120	
Qy	121	ANYVHSGKLGKIYADVGNKTCAGFGSGFYIIDDAQTADGWGVDLLKFDGCGYCDLSLENL	180	
Db	121	ANYVHSGKLGKIYADVGNKTCAGFGSGFYIIDDAQTADGWGVDLLKFDGCGYCDLSLENL	180	
Qy	181	ADGYKMSLALNTRGSRIVSYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Db	181	ADGYKMSLALNTRGSRIVSYSCWPLYMWPFPKPNYTEIRQYCNHWRNFADIDDSWKSIIK	240	
Qy	241	SILDWTSFNQERIVDVAGPGGWNDDPMLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300	
Db	241	SILDWTSFNQERIVDVAGPGGWNDDPMLVGNFGLSNQOVTOMALWAIWAAPLFWMSNDL	300	
Qy	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSLGLAWAVAMINRQEI	360	
Db	301	RHISPOKALLQDKDVIAINQDPLGKQGYQLRQGDNFVWVERPLSLGLAWAVAMINRQEI	360	
Qy	361	GPRSYYTIAVASLKGKGVACNACFITTQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417	
Db	361	GPRSYYTIAVASLKGKGVACNACFITTQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL	417	

RESULT 4

US-10-602-220-16
; Sequence 16, Application US/10602220
; Publication No. US20040023281A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; APPLICANT: POGUE, Gregory P.
; APPLICANT: ERWIN, Robert L.

```

; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KOMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; CURRENT FILING DATE: 2004-05-21
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-16

Query Match          100.0%; Score 2293; DB 5; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  MOLRNPGLHGCAALRFLALYSWDIPGARALDNGLARTPTMGWLHWERFMCNLDCCQEEP 60
DB      1  MOLRNPGLHGCAALRFLALYSWDIPGARALDNGLARTPTMGWLHWERFMCNLDCCQEEP 60

QY      61  DSCISEKLFMEAEMLVMSGKWDAGYEYLCIDD CWMAPQRDSEGR LQADPQRPFGHIGRL 120
DB      61  DSCISEKLFMEAEMLVMSGKWDAGYEYLCIDD CWMAPQRDSEGR LQADPQRPFGHIGRL 120

QY      121  ANYVHSKGLGIYADVGNKTCAGPFGSGGYDDI DAQTADWGV DLLKFDGCGYCD SLENL 180
DB      121  ANYVHSKGLGIYADVGNKTCAGPFGSGGYDDI DAQTADWGV DLLKFDGCGYCD SLENL 180

```

```

QY      181 ADGYKXMSLALNRTGRSYYSCEWPLMYMPFQKPNTYETIRQYCUNHWRNFADIDDSWKSIX 240
Db      181 ADGYKXMSLALNRTGRSYYSCEWPLMYMPFQKPNTYETIRQYCUNHWRNFADIDDSWKSIX 240
QY      241 SILDWTSTFNOERIVDVAGPGGWNDPDLVIIGNFGLSWNQQVTQMALWAJMAAPLPFMSNDL 300
Db      241 SILDWTSTFNOERIVDVAGPGGWNDPDLVIIGNFGLSWNQQVTQMALWAJMAAPLPFMSNDL 300
QY      301 RHISPOAKALLQDKDVIAINODPLGKGQYQLRQGDNFEVWERPLSGLAWAVAMINRQEIG 360
Db      301 RHISPOAKALLQDKDVIAINODPLGKGQYQLRQGDNFEVWERPLSGLAWAVAMINRQEIG 360
QY      361 GPRSYTIIAASLGKVACNPACFIQTLLPVKKRLGFYEWTSLRSHINPTGTVLLOL 417
Db      361 GPRSYTIIAASLGKVACNPACFIQTLLPVKKRLGFYEWTSLRSHINPTGTVLLOL 417

RESULT 6
US-10-984-389-16
; Sequence 16, Application US/10984389
; Publication NO. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-16

```

```
Query Match      100.0%; Score 2293; DB 5; Length 417;
Best Local Similarity 100.0%; Pred. No. 9e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 7
US-09-993-059-12
; Sequence 12, Application US/09993059
; Publication No. US20020086024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993.059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-12

Query Match      100.0%; Score 2293; DB 3; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
```

```
Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 8
US-10-103-327-12
; Sequence 12, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; FILE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103.327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993.059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-12

Query Match      100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60
Db 1 MQLRNPGLHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWRFCMNLDCQEEP 60

Qy 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAMBLMVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120

Qy 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGIGIYADVGNKTCAGFGSGFYDIDAQTFADGWVDLLKFDGCGYCDLSLENL 180

Qy 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240
Db 181 ADGKHMSLALNRTGRSIVVSCWPLYMPPQKPNYTEIRQYCNHWRNFADIDDSWKSIIK 240

Qy 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300
Db 241 SILDWTSEFNOERIVDVAGPGGWNDDPMLVIGNFGLSNQOVTQWALWAIWAAPLFMSNDL 300

Qy 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKDVIATINQDPLGKQGYQLRQGDNEFVWERPLSGLAWAVAMINRQIEG 360

Qy 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417
Db 361 GPRSYYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLOL 417

RESULT 9
US-10-602-219-12
; Sequence 12, Application US/10602219
```

Publication No. US20040016021A1
GENERAL INFORMATION:
APPLICANT: Large Scale Biology Corporation
APPLICANT: Turpen, Thomas H.
APPLICANT: Pogue, Gregory P.
APPLICANT: Erwin, Robert L.
APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSBC-0087-CP09B
CURRENT APPLICATION NUMBER: US/10/602,219
CURRENT FILING DATE: 2003-06-23
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: 09/993,059
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 09/316,572
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: 08/324,003
PRIOR FILING DATE: 1994-10-14
PRIOR APPLICATION NUMBER: 08/176,414
PRIOR FILING DATE: 1993-12-29
PRIOR APPLICATION NUMBER: 07/997,733
PRIOR FILING DATE: 1992-12-30
PRIOR APPLICATION NUMBER: 07/923,692
PRIOR FILING DATE: 1992-07-31
PRIOR APPLICATION NUMBER: 07/600,244
PRIOR FILING DATE: 1990-10-22
PRIOR APPLICATION NUMBER: 07/641,617
PRIOR FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 12
LENGTH: 421
TYPE: PRT
ORGANISM: Homo sapiens
US-10-602-219-12

Query Match 100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLNPGLHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLNPGLHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEIMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEIMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVTSCWPLVYMWFPQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRTGRSIVTSCWPLVYMWFPQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDWTFSNQRIVDVAGPGGNDPDLVIGNFGLSNQVQTMALWAINAALPMSNDL 300
Db 241 SILDWTFSNQRIVDVAGPGGNDPDLVIGNFGLSNQVQTMALWAINAALPMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417
Db 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417

RESULT 10
US-10-602-220-12
Sequence 12, Application US/10602220
Publication No. US20040023281A1
GENERAL INFORMATION:
APPLICANT: Large Scale Biology Corporation
APPLICANT: Turpen, Thomas H.
APPLICANT: Kumagai, Monto H.
APPLICANT: Pogue, Gregory P.
APPLICANT: Erwin, Robert L.
APPLICANT: Grill, Laurence K.
TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
FILE REFERENCE: LSBC-0087-CP07B
CURRENT APPLICATION NUMBER: US/10/602,220
CURRENT FILING DATE: 2003-06-23
PRIOR FILING DATE: 2001-11-13
PRIOR APPLICATION NUMBER: 09/993,059
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 09/626,127
PRIOR FILING DATE: 1999-05-21
PRIOR APPLICATION NUMBER: 08/324,003
PRIOR FILING DATE: 1994-10-14
PRIOR APPLICATION NUMBER: 08/176,414
PRIOR FILING DATE: 1993-12-29
PRIOR APPLICATION NUMBER: 07/997,733
PRIOR FILING DATE: 1992-12-30
PRIOR APPLICATION NUMBER: 08/184,237
PRIOR FILING DATE: 1994-01-19
PRIOR APPLICATION NUMBER: 07/923,692
PRIOR FILING DATE: 1992-07-31
PRIOR APPLICATION NUMBER: 07/600,244
PRIOR FILING DATE: 1990-10-22
PRIOR APPLICATION NUMBER: 07/641,617
PRIOR FILING DATE: 1991-01-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn version 3.2
SEQ ID NO 12
LENGTH: 421
TYPE: PRT
ORGANISM: Homo sapiens
US-10-602-220-12

Query Match 100.0%; Score 2293; DB 4; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MQLNPGLHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Db 1 MQLNPGLHGCALALRFLALVSDIPCARALDNLGLARTPTMGWLHWRFCNLDCCQEEP 60
Qy 61 DSCISEKLFMEAEIMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Db 61 DSCISEKLFMEAEIMVSEGWKADAGYEYLCIDDCWMAPOQDSEGRLOADPQPFPHGIRQL 120
Qy 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTADGWVDLLKFDGCGYCDLSLENL 180
Db 121 ANYVHSGKLGLGIYADVGNKTCAGPFGSGFYDIDAQTADGWVDLLKFDGCGYCDLSLENL 180
Qy 181 ADGKMSLALNRTGRSIVTSCWPLVYMWFPQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Db 181 ADGKMSLALNRTGRSIVTSCWPLVYMWFPQKPNYTEIRQYCNHWNFNADIDDSWSIK 240
Qy 241 SILDWTFSNQRIVDVAGPGGNDPDLVIGNFGLSNQVQTMALWAINAALPMSNDL 300
Db 241 SILDWTFSNQRIVDVAGPGGNDPDLVIGNFGLSNQVQTMALWAINAALPMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQIEG 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLROGDNFEVWERPLSGLAWAVAMINRQIEG 360
Qy 361 GPRSYTIAVSLGKGVACNPACFITQLLPVKRKLGFYEWTSLRSHINPTGTULLQL 417


```
|||||
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 11
US-10-851-388-12
; Sequence 12, Application US/10851388
; Publication No. US20040234516A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/851,388
; PRIOR FILING DATE: 2004-05-21
; PRIOR FILING DATE: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-851-388-12

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCYDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCYDLSLENL 180
Qy 181 ADGKMSLALNRTRGSIVYSCWPLWMPFQKPNYTEIRQYCNHWRNFADIDDSWSKIK 240
Db 181 ADGKMSLALNRTRGSIVYSCWPLWMPFQKPNYTEIRQYCNHWRNFADIDDSWSKIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDPDLVIGNFGLSNNQVQTQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDPDLVIGNFGLSNNQVQTQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 12
US-10-851-388-12
; Sequence 12, Application US/10984389
; Publication No. US20050125859A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/984,389
; PRIOR FILING DATE: 2004-11-08
; CURRENT FILING DATE: 2004-11-08
```

```
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 421
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-389-12

Query Match 100.0%; Score 2293; DB 5; Length 421;
Best Local Similarity 100.0%; Pred. No. 9.1e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Qy 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCYDLSLENL 180
Db 121 ANYVHSGKGLGIYADVGNKTCAGPFGSGFYDIDAQTFADWGVDDLKFGDCCYDLSLENL 180
Qy 181 ADGKMSLALNRTRGSIVYSCWPLWMPFQKPNYTEIRQYCNHWRNFADIDDSWSKIK 240
Db 181 ADGKMSLALNRTRGSIVYSCWPLWMPFQKPNYTEIRQYCNHWRNFADIDDSWSKIK 240
Qy 241 SILDWTSFNQERIVDVAGPGGWNDPDLVIGNFGLSNNQVQTQWALWAIMAAPLFMSNDL 300
Db 241 SILDWTSFNQERIVDVAGPGGWNDPDLVIGNFGLSNNQVQTQWALWAIMAAPLFMSNDL 300
Qy 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Db 301 RHISPOKALLQDKVIAINQDPLGKQGYQLRQDGFVWERPLSGLAWAVAMINRQEI 360
Qy 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417
Db 361 GPRSYTIAVSLGKGVACNACPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLQL 417

RESULT 13
US-09-993-059-18
; Sequence 18, Application US/09993059
; Publication No. US2002008024A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYSOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANSIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/09/993,059
; CURRENT FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-993-059-18

Query Match 100.0%; Score 2293; DB 3; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Db 1 MQLRNPELHGCALALRFLALVSWDIPGARALDNLGLARTPTMGLHWRFCNLDCCOEP 60
Qy 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
Db 61 DSCISEKLFMEAEMLVSEGWKADAGYEYLCIDDCWMAPOQDSRGLQADPQRPFGIRQL 120
```

```
Db 61 DSCISEKLFMEAEIYVADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 120
Qy 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Db 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Qy 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Qy 301 RHISPOKALLQDKVDIATNQDPLGKQYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Db 301 RHISPOKALLQDKVDIATNQDPLGKQYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSHINPTGTVLQL 417

RESULT 14
US-10-103-327-18
; Sequence 18, Application US/10103327
; Publication No. US20030106095A1
; GENERAL INFORMATION:
; APPLICANT: GARGER, Stephen A.
; APPLICANT: TURPEN, Thomas H.
; APPLICANT: KUMAGAI, Monto H.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN
; TITLE OF INVENTION: PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: 008010087CPUS06
; CURRENT APPLICATION NUMBER: US/10/103,327
; CURRENT FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US/09/993,059
; PRIOR FILING DATE: 2001-11-13
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-103-327-18

Query Match 100.0%; Score 2293; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFERNLDCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFERNLDCQEEP 60
Qy 61 DSCISEKLFMEAEIYVADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 120
Db 61 DSCISEKLFMEAEIYVADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 120
Qy 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Db 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Qy 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Qy 301 RHISPOKALLQDKVDIATNQDPLGKQYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
```

```
Db 301 RHISPOKALLQDKVDIATNQDPLGKQYQLROGDNFEVWERPLSGLAWAVAMINROEIG 360
Qy 361 GPRSYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSHINPTGTVLQL 417
Db 361 GPRSYTIAVASLGKGVACNACPFITQLLPVKRKLGFYEWTSRLSHINPTGTVLQL 417

RESULT 15
US-10-602-219-18
; Sequence 18, Application US/10602219
; Publication No. US20040016021A1
; GENERAL INFORMATION:
; APPLICANT: Large Scale Biology Corporation
; APPLICANT: Turpen, Thomas H.
; APPLICANT: Pogue, Gregory P.
; APPLICANT: Erwin, Robert L.
; APPLICANT: Grill, Laurence K.
; TITLE OF INVENTION: PRODUCTION OF LYOSOMAL ENZYMES IN PLANTS BY TRANIENT EXPRESSION
; FILE REFERENCE: LSBC-0087-CP09B
; CURRENT APPLICATION NUMBER: US/10/602,219
; CURRENT FILING DATE: 2003-06-23
; PRIOR APPLICATION NUMBER: 09/993,059
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 09/626,127
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 09/316,572
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 08/324,003
; PRIOR FILING DATE: 1994-10-14
; PRIOR APPLICATION NUMBER: 08/176,414
; PRIOR FILING DATE: 1993-12-29
; PRIOR APPLICATION NUMBER: 07/997,733
; PRIOR FILING DATE: 1992-12-30
; PRIOR APPLICATION NUMBER: 08/184,237
; PRIOR FILING DATE: 1994-01-19
; PRIOR APPLICATION NUMBER: 07/923,692
; PRIOR FILING DATE: 1992-07-31
; PRIOR APPLICATION NUMBER: 07/600,244
; PRIOR FILING DATE: 1990-10-22
; PRIOR APPLICATION NUMBER: 07/641,617
; PRIOR FILING DATE: 1991-01-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 18
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-602-219-18

Query Match 100.0%; Score 2293; DB 4; Length 423;
Best Local Similarity 100.0%; Pred. No. 9.2e-220;
Matches 417; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFERNLDCQEEP 60
Db 1 MOLRNPELHGCALALRFLALVSWDIPGARALDNGLARTPTMGWLHWFERNLDCQEEP 60
Qy 61 DSCISEKLFMEAEIYVADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 120
Db 61 DSCISEKLFMEAEIYVADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 120
Qy 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Db 121 ANYVHSGKLGIGIYADGNKTCAGFGSGFYDIDAQTADGWVDLLKFDGCGYCDSENL 180
Qy 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Db 181 ADGKXMSLALNRTGRSIVYSCWPLVYMWPPQKNTYTIROYCNHWNFPADIDDSWSIK 240
Qy 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
Db 241 SILDWTSFNOERIVDVAGPGGNDPDMVLVGNFGLSNQVQVQVQVQVQVQVQVQVQVQV 300
```

Qy	301	RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFEVWERPLSGIAVAVAMINROEIG	360
Db	301	RHISPOAKALLQDKDVIAINQDPLGKQGYQLROGDNFEVWERPLSGIAVAVAMINROEIG	360
Qy	361	GPRSYTTIIVASLKGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL	417
Db	361	GPRSYTTIIVASLKGKGVACNPFITQLLPVKRKLGFYEWTSRLRSHINPTGTVLLQL	417

Search completed: January 1, 2006, 00:29:24
Job time : 124.703 secs

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

This Page Blank (uspto)